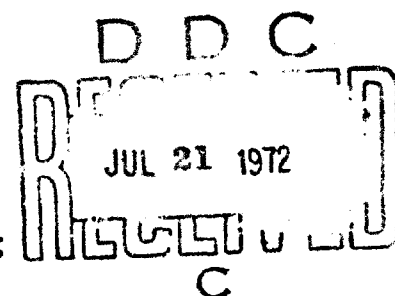ELECTRICAL CONDUCTION IN DISCONTINUOUS FILMS ON AN INSULATING SUBSTRATE

Joseph E. Dryer

The Ohio State University
Department of Metallurgical Engineering

D D C

JUL 21 1972

C

TECHNICAL REPORT AFML-TR-72-68

March 1972

Approved for public release; distribution unlimited.

AIR FORCE MATERIALS LABORATORY
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely related Government procurement operation, the United States Government thereby incurs no responsibility nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Copies of this report should not be returned unless return is required by security considerations, contractual obligations, or notice on a specific document.

DOCUMENT CONTROL DATA - R & D

Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
The Ohio State University Research Foundation 1314 Kinnear Road Columbus, Ohio 43202	Unclassified
	2b. GROUP

3. REPORT TITLE

ELECTRICAL CONDUCTION IN DISCONTINUOUS FILMS ON AN INSULATING SUBSTRATE

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Final Report   4/30/69 to 4/15/72

5. AUTHOR(S) (First name, middle initial, last name)

Joseph E. Dryer

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
March 24, 1972	113	78

8a. CONTRACT OR GRANT NO	9a. ORIGINATOR'S REPORT NUMBER(S)
F33615-69-C-1026	
b. PROJECT NO 7353	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) AFML-TR-72-68
d.	

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Air Force Materials Laboratory Air Force Systems Command Wright-Patterson AFB, Ohio 45433

13. ABSTRACT

Experimental and theoretical investigation of the electrical properties of thin films during the nucleation stage of film growth was undertaken. It was found that electrostatic forces in the film can significantly influence the film resistance, nucleation rate and island size. Electrode contact effects were also found to have a large effect in measurements taken across narrow films. Experimentally it was found that a number of films investigated exhibited a reversible switching phenomena, wherein they underwent a substantial increase in conductivity when subjected to large voltages. This high conductivity then persisted to lower voltages and subsequently the film could be converted to its original condition. Previous theories of conduction in discontinuous films were considered and modified to include electrode carrier injection and substrate dielectric polarization.

KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Island Growth						
Thin Films						
Nucleation						
Switching						
Discontinuous Films						

ELECTRICAL CONDUCTION IN DISCONTINUOUS FILMS ON AN INSULATING SUBSTRATE

Joseph E. Dryer

ic

FOREWORD

This report was prepared by the Department of Metallurgical Engineering, The Ohio State University, Columbus, Ohio. The principal investigators for this project were Dr. John P. Hirth and Dr. Rudolph Speiser under Contract F33615-69-C-1026. This contract was initiated under Project 7353, "Research on Intrinsic Surface Properties and Their Relation to the Mechanical Behavior of Metals and Alloys," during the period November, 1968 through April, 1972. This Final Report is the Ph.D. dissertation of Dr. Joseph E. Dryer based upon the research conducted under this contract. This work was administered by the Advanced Metallurgical Studies Branch of the Metals and Ceramics Division, Air Force Materials Laboratory, Wright-Patterson Air Force Base, Ohio, under the direction of Dr. Harold L. Gegel.

This report was submitted by the author in March 1972.

This technical report has been reviewed and approved.

C.T. LYNCH
Chief, Advanced Metallurgical
            Studies Branch
Metals and Ceramics Division
Air Force Materials Laboratory

ii

ABSTRACT

Experimental and theoretical investigation of the electrical properties of thin films during the nucleation stage of film growth were undertaken. It was found that electrostatic forces in the film can significantly influence the film resistance, nucleation rate and island size. Electrode contact effects were also found to have a large effect in measurements taken across narrow films. Experimentally it was found that a number of films investigated exhibited a reversible switching phenomena wherein they underwent a substantial increase in conductivity when subjected to large voltages. This high conductivity then persisted to lower voltages and subsequently the film could be converted to its original condition. Previous theories of conduction in discontinuous films were considered and modified to include electrode carrier injection and substrate dielectric polarizaticn.

TABLE OF CONTENTS

Preceding page blank

LIST OF FIGURES

LIST OF FIGURES - Continued

SECTION I

INTRODUCTION

This paper contains a theoretical and experimental investigation of some of the unique resistive properties of very thin gold films on insulating substrates. When the films are too thin to support a continuous structure they assume the form of a network of "islands" separated by the substrate. There is, therefore, no continuous metal path for current flow from one side of the film to the other. Voltages applied across such films, however, result in a significant current, the magnitude of which depends strongly on the amount of material deposited and the deposition conditions.

One uncommon feature of discontinuous films is a static resistance which can show a marked decrease with increasing temperature. Experimental data can be evaluated to give a thermal activation energy of between 0.001 and 1 eV, with the higher energies corresponding to very thin films of high melting point metals and the smaller energies corresponding to nearly continuous films. This temperature dependence is opposite the positive temperature coefficient of resistance of bulk metals and thicker, continuous films.

Another singular feature of discontinuous films is a highly nonlinear static resistance. The logarithm of the resistance is often proportional to the square root of the applied voltage. Discontinuous films show an increase in resistance with room-temperature and high-temperature anneals.

Experimental data will be presented which verify the above-mentioned phenomena in discontinuous films. This data, as well as new phenomena found in these films, will be shown to be consistent with a conduction mechanism which will be proposed.

Other features of conduction in discontinuous films which have been reported in the literature, but which will not be considered in this paper, are a lowering of the work function below that of the bulk metal,[1] a decrease in resistance with increasing frequency[2] and photoluminescence when current is passed through the films.[3] Many of the optical, structural, and magnetic properties of these films also deviate markedly from those properties expected from an extrapolation of the bulk and thick film behavior.

The discontinuous, island structure of very thin films is a consequence of the large surface areas present in very thin films. The driving forces for the island geometry is discussed in Appendix I.

In this report the energy barrier for charge transport in discontinuous films is assumed to be at least partly due to the electrostatic energy involved in the localization of a charge. An investigation of the form of this activation energy for various island geometries shows that the energy is inversely proportional to the longest island dimension. For a static charge on an island the energy is also inversely proportional to the substrate low frequency dielectric constant. It is shown that the use of the low frequency dielectric constant modifies the form of the carrier transport equations in the case of a fast transport, such as tunneling or thermal activation over a high energy barrier. The modified equations, taking the dielectric relaxation time into account, do not agree well with the experimental data. A carrier transport by short carrier "hopping" between trapping levels in the substrate is proposed.

With this mechanism for carrier transport the activation energy is purely electrostatic. Carrier creation by electrode injection would, therefore, require a much lower energy than carrier creation in the bulk of the film. Once a carrier is created, this carrier can then move with the addition of the much smaller energy because of the electrostatic attraction of a charge to a neutral island. The reduction of the resistance in such films with increasing voltage is attributed to the field-induced lowering of the barrier seen by an electron leaving an electrode.

The contact injection model is used to explain the experimentally observed voltage effects on the film resistance, the time dependence of the resistance change during film annealing, and the resistance change caused by variation in the spacing of the electrodes between which the film was deposited.

A new phenomena, the film thickness dependent "breakdown" of the film resistance, is discussed. When a sufficiently large electric field is applied, several of the films underwent an abrupt decrease of several orders of magnitude in the film resistance. This lower resistance then persisted to very low applied voltages. The films could subsequently revert to a high resistance state, and be cycled in and out of the low resistance state many times.

The theoretical model for conduction in thin, discontinuous films allows a better understanding of this type of film. Because of the continuously variable temperature coefficient of resistance and the large sheet resistances possible, these films have potential application in thin film microelectronics. The interaction between charged and uncharged islands, and the stresses on an island due to a charge on that island modify the driving forces in thin film nucleation and island agglomeration. Some experimental evidence of such effects is discussed.

SECTION II

REVIEW OF THE LITERATURE

The unique resistive properties of discontinuous films have been long known and have been discussed theoretically in a number of articles. All theories must explain the resistance which has a log which is proportional to the inverse of temperature, with the constant of proportionality continuously decreasing with increasing average film thickness. At sufficiently high applied fields the films all show a resistance which is nonlinear, the log of the resistance decreasing by a factor proportionally to the square root of voltage.

One of the most widely utilized theories is that originally developed by Neugebauer, (References 19, 20, 21). In this approach it is assumed that the placing of a charge on an island requires an amount of energy equal to that total electrostatic energy of the field due to that charge. The number of islands with charges can then be expressed using Maxwell-Boltzmann statistics. It was assumed that the charge transport giving rise to current is caused by tunneling of these charges from one island to the next. The tunneling is expressed by a WBK approximation, with no energy difference between the initial and final states. The resistance is proportional to the product of the activated supply function and the tunneling transport function, with the temperature dependence of the form $\exp(-E_a/kT)$, where E_a is the electrostatic activation energy. A Poole-Frenkel type of field lowering of the electrostatic energy was assumed to account for the observed voltage dependence of the current.

Hartman[22] has proposed a modification of the tunneling phenomenon involving the fact that the energy is no longer a quasi-continuous function of crystal momentum for particles on the atomic scale, but rather exists in discrete levels with non-negligible energy gaps between the levels. Tunneling between islands must be preceded by thermal activation of a carrier over this energy gap to the first activated level in order to tunnel to an unoccupied site in the adjacent island. This theory predicts a resistance with the same temperature dependence as the Neugebauer theory.

Herman and Rhodin[23] have suggested that the interaction between the islands and the substrate affect the trap levels under the islands. Phonon-assisted tunneling of carriers from occupied trapping levels under the island to unoccupied levels between islands yields results that agree with the observed temperature dependence. This theory extends the earlier work of Wei[24] and combines it with Neugebauer's tunneling of activated charges. Wei proposed that the islands act as donor states contributing carriers to the conduction band of the insulating substrate.

3

Historically, thermionic emission was the first phenomenon applied to explain the temperature effects observed in thin film resistance. The small values often observed for the activation energy appear to be incompatible with either the work function or the normal contact potential between the substrate and the metal. Minn[25] proposed that the work function was lowered by the image force on the electron. This proposal was further developed by van Steensel,[26] who carried out a number of interesting experiments. The films were grown on ferroelectric barium titanate, and the substrate was cycled through its Curie temperature causing a large change in the bulk dielectric constant with little variation in the resistance of the discontinuous films. It was claimed by van Steensel[26] that the barrier lowering caused by overlapping of the image charges between islands was sufficient to explain the observed low activation energies.

Hill[27],[28] has assumed that tunneling is the dominant current transport mechanism, and that a number of trapping sites exist in the substrate between islands; the tunneling is by means of these traps. By applying an electric field perpendicular to the plane of the film, Hill was able to modulate the conductivity of his discontinuous films. He assumed that this conductivity change was a result of the alkali ion migration to and from the surface under the influence of the field, modulating the number of available traps.

Phonon-assisted tunneling from island to island has been recently proposed by Hofer and Fromm[29] to account for the observed data. This approach is basically an extension of the work of Herman and Rhodin.[23]

ELECTROSTATIC FIELDS

With the possible exception of Wei's theory of discontinuous film conduction[24] in which the islands act as donor centers contributing carriers to the conduction band of the substrate, treatments of the problem of current flow in discontinuous films assumed that there are charges present on some of the islands. Whenever an uncompensated charge is localized some energy is concentrated in the electrostatic field of that charge. It is, therefore, an essential feature of all the theories involving the localization of a charge on an island that some activation process must be invoked to supply the electrostatic energy. This section will examine the form of the electrostatic energy and some of the consequences of the charge localization.

The most extensive discussion of the origin and nature of the activation energy is given in an article by R. M. Hill.[27] Hill assumes the islands are spherical and completely immersed in the dielectric substrate for ease of calculation. It has been long recognized[30] that under such conditions an island charged by an amount, e, will have its potential energy raised with respect to zero potential energy in an infinitely large island by an amount:

$$E_0 = \frac{e^2}{4\pi \epsilon_\ell a} \tag{1}$$

where a is the radius of the island, and ϵ_ℓ is the low frequency dielectric constant of the substrate. The potential away from a charged spherical conductor in the absence of any other conductors drops off inversely with the distance from the center of the sphere as shown in Fig. 1. Hill considered a row of such spheres, one of which was charged, and considered the potential on another particle a distance, D, away to be $e^2/4\pi \epsilon_\ell D$, so the potential difference between the two spheres, defined as δE_0, is given by

$$\delta E_0 = \frac{e^2}{4\pi \epsilon_\ell} \left\{ \frac{D-a}{aD} \right\} \tag{2}$$

This energy is considered to be the work done in transporting the charge from the first to the second particle, and, therefore, to be the activation energy of the film. It should be noted that Hill considered charge transport by tunneling. This energy is, therefore, not the energy barrier the electron must surmount, but rather the energy difference between the system before and after the charge transport. In the presence of an applied field along the row of islands, Hill considered the situation
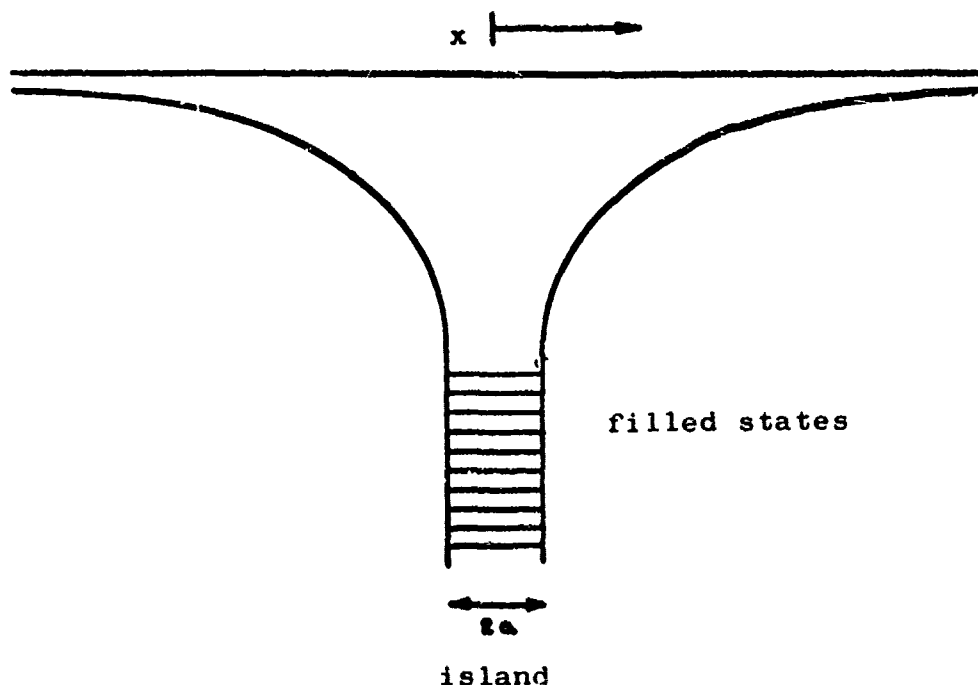
Fig. 1 - Potential in the Vicinity of an Island
With No External Electric Field

as described in Fig. 2. The potential energy is given by

$$E = \frac{e^2}{4\pi\epsilon_\ell x} + e\mathcal{E}x + C \tag{3}$$

where $\mathcal{E}$ is the applied field in the x direction (x = 0 at the center of the charged island) and C is a constant. This expression has a maximum which can be shown to be an amount δE below the potential energy for no applied field. The energy corresponding to this barrier reduction is δE and is given by

$$\delta E = -2\left(\frac{e^3}{4\pi\epsilon_\ell}\mathcal{E}\right)^{1/2} \tag{4}$$

δE is taken to be the reduction in the activation energy for large fields. When the field is small the reduction in the activation energy is given by the voltage drop to the next island, resulting in

6
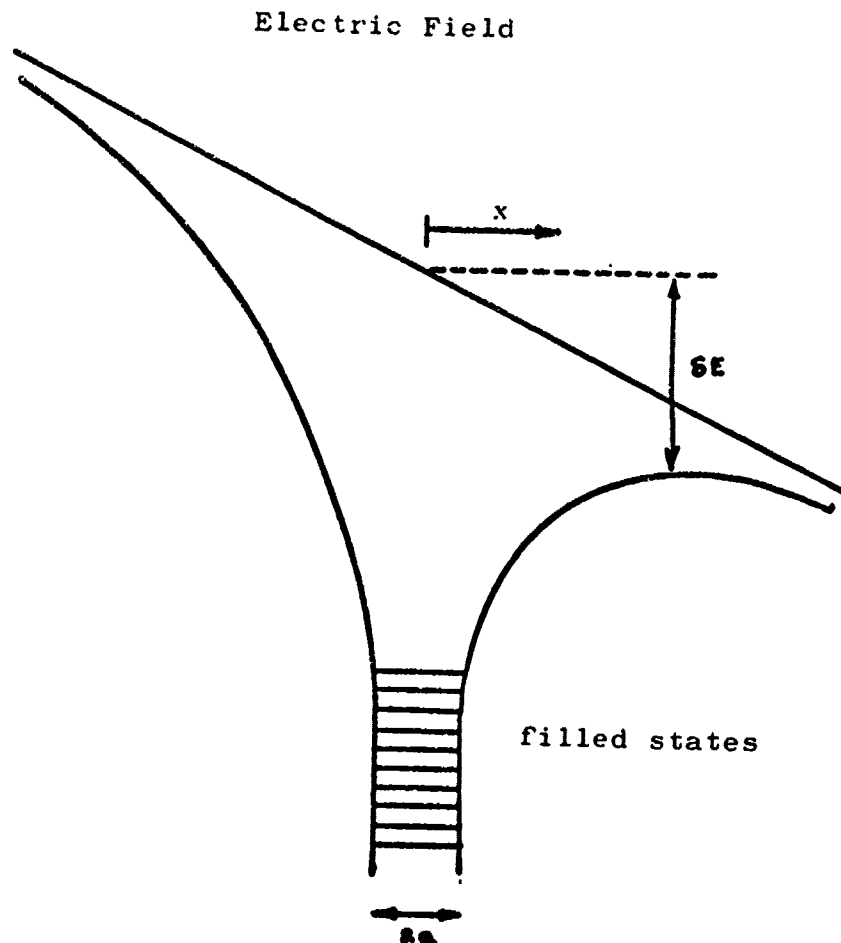
Electric Field

Fig. 2 - Potential in the Vicinity of an Island
With An External Electric Field

$$\delta E = \delta E_0 - e \zeta D \tag{5}$$

where δE_0 is given by Eq. (2). None of these expressions is precisely correct, and an analysis of the reasons for the mistakes will be a useful introduction to a consideration of the electrostatic forces in discontinuous thin films.

In the following discussion, the electrostatic energy of a particle will be the energy in the electrostatic field surrounding the particle because of a charge on the particle, including the interactions of that charge with neighboring conductors. The energy could be calculated by determining the electric field everywhere because of that charge and determining the electrostatic energy, E_c, as

7

$$E_c = \frac{1}{2} \int_{\substack{all \\ space}} \epsilon \, \mathcal{E}^2 \, d\overline{v} \qquad (6)$$

An alternative,[31] but entirely equivalent, method of determining E_c is to consider the N initially charged conductors in the system with the ith conductor having a charge Q_i. A charge on any conductor will affect the potential of all other conductors. Therefore, if V_i is the potential of the ith conductor, we may write

$$
\begin{aligned}
V_1 &= S_{11}Q_1 + S_{21}Q_2 + \cdots + S_{n1}Q_n \\
V_2 &= S_{12}Q_1 + S_{22}Q_2 + \cdots + S_{n2}Q_n \\
&\cdots \\
V_n &= S_{1n}Q_1 + S_{2n}Q_2 + \cdots + S_{nn}Q_n .
\end{aligned}
\qquad (7)
$$

The coefficients S_{ij} are called the coefficients of potential or the mutual susceptances. Green's reciprocity theorem can be used to show that $S_{ij} = S_{ji}$. This set of equations can also be solved for the charge on the conductors in terms of the potentials of all the conductors as

$$
\begin{aligned}
Q_1 &= C_{11}V_1 + C_{21}V_2 + \cdots + C_{n1}V_n \\
Q_2 &= C_{12}V_1 + C_{22}V_2 + \cdots + C_{n2}V_n \\
&\cdots \\
Q_n &= C_{1n}V_1 + C_{2n}V_2 + \cdots + C_{nn}V_n .
\end{aligned}
\qquad (8)
$$

The quantity C_{rr} is called the coefficient of capacitance or the self capacitance, and C_{ij} is called the coefficient of induction or the mutual capacitance. Since the [S] matrix and the [C] matrix are inverses of each other, we know that S_{ij} is the minor of C_{ij} divided by the determinant of the C matrix. Since the energy of a system is given by

$$E_c = \frac{1}{2} \sum_{n=1}^{n} V_i Q_i \qquad (9)$$

which is equivalent to Eq. (6), E_c can be written as

$$E_c = \frac{1}{2} \left(C_{11}V_1^2 + C_{12}V_1V_2 + C_{22}V_2^2 + \cdots \right) \qquad (10)$$

or

$$E_c = \frac{1}{2} \left(S_{11}Q_1^2 + 2 S_{12}Q_1 Q_2 + S_{22}Q_2^2 + \cdots \right) \tag{11}$$

If we have charge on only one conductor, Eq. (11) is the easiest form to use since as $Q_2 = Q_3 = \cdots = 0$, there is only one term. Often, however, it is easier to calculate the capacitive coefficients and then invert the matrix to find the proper coefficient of potential to use in Eq. (11).

Two further points should be made. In the case of an isolated conductor of capacitance C (defined as the charge divided by the potential with respect to a zero at infinity caused by that charge) Eq. (11) becomes

$$E_c = \frac{1}{2} \frac{Q^2}{C} \tag{12}$$

Secondly, if the charge leaves a conductor and exists as a point charge at a position between the conductors, the electrostatic potential potential becomes infinite because of the self-energy of the electron. This problem will be discussed later; presently the charge will be assumed always to reside on a conductor.

For a spherical conductor, if we consider the electric field crossing a spherical surface of radius, r, centered on the sphere with charge, e, Gauss's law gives the electric field as

$$\mathcal{E} = \frac{e}{4\pi\epsilon_\ell r^2} \tag{13}$$

Since

$$\dot{\mathcal{E}} = -\nabla V \quad ,$$

$$V = \frac{e}{4\pi\epsilon_\ell r} + A \quad . \tag{14}$$

If V goes to zero at infinity, then A = 0, and the surface of the sphere (radius a) is at a potential

$$V_o = \frac{e}{4\pi\epsilon_\ell a} \tag{15}$$

9

Therefore, the capacitance of the sphere is given by

$$C = 4\pi \epsilon_\rho a \qquad (16)$$

and the electrostatic energy of the sphere, given by Eq. (12) is

$$E_c = \frac{e^2}{8\pi \epsilon_\rho a} \cdot \qquad (17)$$

It is seen that this is half the value implied by Eq. (1). This discrepancy arises from the assumed source of the charge that is placed on the island. If that charge comes from another island, there are two islands, one charged +e, the other charged -e. Therefore each island has an individual electrostatic energy given by Eq. (17), and the total energy required to create the pair is given by Eq. (1). Equation 17 expresses the value of energy that should have been used as a starting point in Reference 27.

Although contrary to the statement of Hill,[27] his description most accurately represents the transfer of charge from an initially neutral island to an adjacent island. It can be seen that there are two problems to be considered: (1) the creation of a charged island and (2) the movement of the charge once it is isolated on an island. In any detailed attempt to calculate the energy barriers and activation energies involved in discontinuous film resistances, the differences between these two cases must be included.

As discussed previously, the electrostatic energy is calculated assuming the charge is at rest on a conductor. If the islands are equally spaced and of an equal size, by symmetry, the electrostatic energy after a charge has moved to the next island, and the dielectric has had time to adjust, is the same as the electrostatic energy before the charge moved. Problems associated with the relaxation of the dielectric are discussed later. In this sense, E is not the work that must be done in transporting the charge from the first to the second particle. Figure 1 is misleading because it considers the potential due to a charge on an island and assumes this potential remains the same once the charge has left the island. This consideration of the system remaining in the same energy state after a mobile charge has moved from one island to the next island was recognized by Neugebauer and was one of the basic features of his treatment.

Equation 4 represents the maximum in the potential barrier due to the charge. Even if this barrier remained in the same form as the charge transverses it, the energy difference at the next island a fixed distance, D, away would have the form of Eq. (5) irregardless of the field strength. The ϵ dependence of Eq. (4) is caused by the movement of the potential

10

maximum with the electric field. If a tunneling transport is assumed, the barrier height cannot be called an activation energy in the normal sense.

Returning to the spherical conductor, it will be illustrative to calculate, by a variation of the method used by Faraday,[33] the stress on the surface of the sphere due to the charge enclosed. If we consider an infinitesimal element of volume oriented in an electric field so that two surfaces are normal to the field and the other surfaces are parallel to the electric field we can evaluate the forces on the two normal surfaces due to the electric field. If one of the surfaces is held fixed, the normal stress on the other surface is given by

$$P = - \frac{dE}{dV}$$

where V is the volume of the element and E is the energy of the element. The electrostatic energy in the element, if the element is small enough to have a constant $\mathcal{E}$ field through it, is

$$E = \frac{1}{2} \epsilon \, \mathcal{E}^2 \, V \tag{18}$$

giving a stress because of the electrostatic forces of magnitude

$$|P| = \frac{1}{2} \epsilon \, \mathcal{E}^2 \tag{19}$$

At the surface of a metal, if the density of surface charge is σ, we know from Gauss's theorem that

$$\mathcal{E} = \frac{\sigma}{\epsilon} \tag{20}$$

From Eqs. (19) and (20) we see that there will be an outward stress on the conduction surface given by

$$P = \frac{\sigma^2}{2\epsilon} \tag{21}$$

This force is independent of the sign of the charge on the surface.

In the case of the spherical conductor the charge is uniformly distributed over the surface, so that if e is the total charge on the sphere,

$$\sigma = \frac{e}{4\pi a^2}$$

(22)

This gives a stress on the surface of the sphere of

$$P_{sph} = \frac{e^2}{32\pi^2 \epsilon a^4}$$

(23)

For e equal to one electronic charge, this reduces to

$$P_{sph} = \frac{1.02}{\epsilon_r a^4} \times 10^{10} \frac{newtons}{m^2} = \frac{1.02}{\epsilon_r a^4} \times 10^{11} \frac{dynes}{cm^2}$$

(24)

where ϵ_r is the relative low frequency dielectric constant and a is the radius of the sphere in angstroms.

For the next higher order of complexity, consider two spherical conductors as shown in Fig. 3a. To calculate the capacitive coefficients as given by Eq. (8) for this system, it can be assumed that sphere 1 is at unit potential by placing a charge $4\pi\epsilon a$ at its center. If sphere 2 is at ground potential, a charge will be induced of magnitude $-\frac{4\pi\epsilon a^2}{D}$ at a distance $\frac{a^2}{D}$ to the left of the center of sphere 2 by the method of charge images in spheres. This charge in sphere 2 will have its image inside sphere 1, which in turn will have its image inside sphere 2, etc. We can then add up the charges inside sphere 1, which will be C_{11} since sphere 1 was held at unit potential. By symmetry this will also be C_{22}. The total charge induced in sphere 2 is $C_{12}(= C_{21})$. This can be shown to give[32]

$$C_{11} = C_{22} = 4\pi\epsilon a \sinh\beta \sum_{n=1}^{\infty} csch\,(2n-1)\beta \quad ,$$

(25)

$$C_{12} = C_{21} = -4\pi\epsilon a \sinh\beta \sum_{n=1}^{\infty} csch\,2n\beta \quad ,$$

(26)

$$\beta = \pm \cosh^{-1}\left[\frac{D^2 - 2a^2}{2a^2}\right] \quad .$$

(27)

(a) Two Charged Spheres

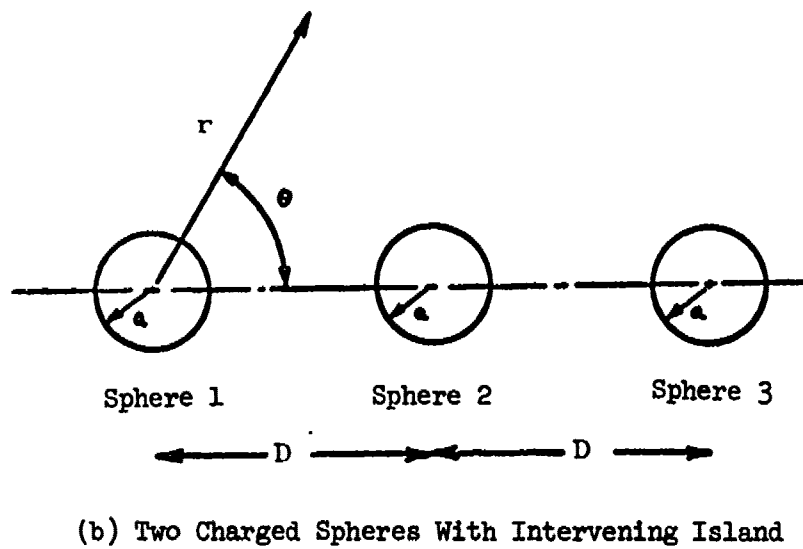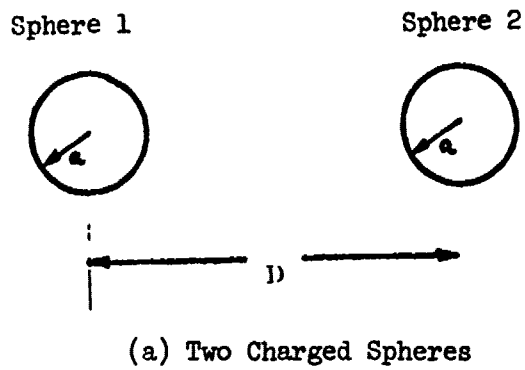(b) Two Charged Spheres With Intervening Island

Fig. 3 - Multiple Charged Spheres

From these equations the coefficients of potential can be calculated
and the electrostatic energy can be determined from Eq. (11) for any
charge distribution on the two spheres. Carrying out these calculations
it can be seen that if charge is on one sphere, the electrostatic energy
is given by

$$E_c' = E_c \; \frac{1}{\sinh\beta} \; \frac{\sum\limits_{n=1}^{\infty} \operatorname{csch}(2n-1)\beta}{\left[\sum\limits_{n=1}^{\infty} \operatorname{csch}(2n-1)\beta\right]^2 - \left[\sum\limits_{n=1}^{\infty} \operatorname{csch} 2n\beta\right]^2} \qquad (28)$$

where E_c is given by Eq. (17), and is the electrostatic energy of the
charged sphere in the absence of the uncharged sphere. If one sphere
is charged positively and the other is charged negatively, as when
charge is created, then the total energy is

$$E_{c_{TOT}} = 2E_c' - E_c' \; \frac{2\sum\limits_{n=1}^{\infty} \operatorname{csch} 2n\beta}{\left[\sum\limits_{n=1}^{\infty} \operatorname{csch}(2n-1)\beta\right]^2 - \left[\sum\limits_{n=1}^{\infty} \operatorname{csch} 2n\beta\right]^2} \frac{1}{\sinh\beta} \qquad (29)$$

The last term represents the reduction in the total electrostatic energy
due to the interaction between the charged islands. If, for example,
$D = \sqrt{6}\, a$, then Eq. (28) predicts only a 5% decrease in the electrostatic
energy of one charged sphere, while Eq. (29) predicts a further decrease
to about 0.543 of the value given by twice Eq. (17). It should be noted
that Eq. (2) predicts a total energy of 0.59 times the value given by
Eq. (17).

The calculation can be extended to the case of three spheres
arranged as shown in Fig. 3b. If a charge of plus and minus e resides
on the two outer spheres, this would approximate the situation of Fig.
3a after the charge has made one more hop. The calculation of the self
and mutual capacitances could be carried out by the use of an infinite
series of images. A much more rapid approach to approximate the total
electrostatic energy would be to calculate the energy of the two outer
spheres with no inner sphere present by the use of Eqs. (28) and (29),
and then calculate the additonal energy given up as the middle sphere
is drawn into position from an infinite distance away. A conservative
estimate of this energy can be made by assuming that a charge, e, a dis-
tance, r, away from an uncharged sphere of radius, a, induces image
charges in the sphere which can be approximated for large r by a dipole
of magnitude $a^3 e/r^2$ located on the center of the sphere. Since the
force on a dipole $\overline{M}$ is $(\overline{M}\cdot\nabla)\overline{\mathcal{E}}$, where $\overline{\mathcal{E}}$ is the magnitude of the elec-
tric field produced by the charge, it is found that the force on the

sphere is

$$|F| = \frac{a^3 e^2}{2\pi \epsilon r^5} \tag{30}$$

This is a good approximation of the force of attraction between a spherical conductor located a distance, r, away if r is much larger than a. For smaller values of r the force is actually larger than given by Eq. (30). Therefore, if the center sphere is located initially at a point above the midpoint between spheres 1 and 3, the resultant downward force on this sphere caused by the charges ± e on spheres 1 and 3 is given by

$$F_{down} = \frac{a^3 e^2}{4\pi\epsilon_0} \frac{\cos^5\theta}{(D/2)^5} \sin\theta \tag{31}$$

where θ is defined in Fig. 3b. The total energy extracted from the electrostatic field in bringing this sphere from an infinite distance away to the position shown in Fig. 3b is given by

$$\Delta E = \int_0^\infty \frac{a^3 e^2}{4\pi\epsilon} \frac{\cos^5\theta}{(D/2)^5} \sin\theta \frac{D}{2} d(\tan\theta) = \frac{e^2}{4\pi\epsilon a} \frac{8a^4}{D^4} \tag{32}$$

This value, subtracted from that of Eq. (29) will then give a lower limit to the electrostatic energy of the charged spheres in this configuration.

An example, if the ratio of D to a is $\sqrt{6}$ as was considered before, Eq. (2) would predict an activation energy of about 0.795 times twice Eq. (17), in close agreement with the value of 0.8 obtained from Eq. (29). Equation (32), however, predicts a further lowering of the energy by a factor of 0.22 to a total of 0.68 of Eq. (17) due to the presence of the uncharged sphere between the two charged spheres. This energy lowering would become more significant as D became smaller with respect to a, both because of the power relationship in Eq. (32) and the inaccuracy in the derivation of that equation.

Although the derivation could be carried out to more spheres, the principal point has been made. By excluding the effect of intermediate islands lowering the potential illustrated in Fig. 1a, Eq. (2) represents an activation energy which increases much too rapidly toward the values for infinite charge separation. The more extended field encountered in practice has implications in the field lowering of the barrier which is discussed later.

In actual films, however, the conductors are not spherical and immersed in a dielectric. A more realistic model would be a disk-shaped or hemispherical co ductor with the dielectric on one side and vacuum on the other side a shown in Fig. 4.
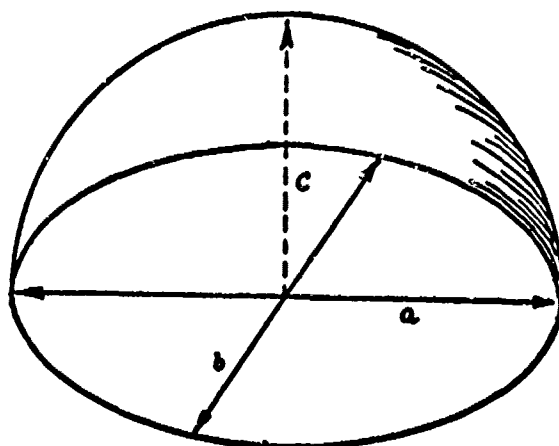
Fig. 4 - Disc-Shaped Island

The most quoted discussion on nonspherical effects is in the article by R. M. Hill.[27] The following is taken from his article.

> As the electron microscope provides information about the diameter of a particle in the plane of the substrate the activation energies will be expressed in terms of the equivalent spherical particle. It is assumed that the cross-section in the substrate plane remains circular. A closer approximation to the real shape is probably that of a hemispherical cap. For single particles of this form the coulombic potential is $1/2$ E_c, and if the cap flattens to an infinitely thin disk it becomes $\pi/2$ E_c. The most useful approximation is that of an ellipsoidal particle. If ℓ is the eccentricity and δ is the ratio of the axis perpendicular to the substrate to that in the substrate plane then the coulombic potential is

$$\frac{E_c}{2} \ln\left(\frac{1+\ell}{1-\ell}\right)(\ell\delta)^{-1} \tag{33}$$

If a cap-shaped island interface with the substrate is a circle of radius r, then the island can be considered to be a portion of a sphere of radius a. As was done by Hill, it will temporarily be assumed that all conductors considered will be immersed in the dielectric. Let E_c'

16

be the initial electrostatic energy of the cap-shaped island and E_c (given by Eq. (17)) be the electrostatic energy of the imaginary conducting sphere of which the cap-shaped island can be considered a part. Consider a conductor shaped so as to form a sphere of radius a when fitted to the cap of radius r. If the charge is initially in the cap, and the other conductor is located a large distance away and uncharged, this conductor will experience an attractive force toward the charge caused by induced polarization of its fixed and mobile charge. This conductor will then be pulled to the cap through a resistive medium, doing work which is extracted from the electrostatic energy of the charged cap. When the parts join together to form a sphere, the electrostatic energy will be E_c which must be less than E_c', giving

$$E_c' > \frac{e^{\cdot\cdot}}{8\pi\epsilon_\ell a} \tag{34}$$

Thus, using the radius of the base of a disk-shaped cap, in the electrostatic energy equation for a sphere will always give an energy that is too low. Note this is in contradiction with Hill's result that the electrostatic energy of a hemispherical cap is one-half the electrostatic energy of an equivalent sphere.

If Laplace's equation is written in an ellipsoidal coordinate system[14] (ξ, η, ζ),

$$\nabla^2\psi = \frac{4}{(\xi-\eta)(\xi-\zeta)(\eta-\zeta)} \left[(\eta-\zeta)R_\xi \frac{\partial}{\partial\xi}\left(R_\xi \frac{\partial\psi}{\partial\xi}\right)\right.$$
$$\left. + (\zeta-\xi)R\eta \frac{\partial}{\partial\eta}\left(R\eta \frac{\partial\psi}{\partial\eta}\right) + (\xi-\eta)R_\zeta \frac{\partial}{\partial\zeta}\left(R_\zeta \frac{\partial\psi}{\partial\zeta}\right)\right] = 0 \tag{35}$$

where

$$R_s = \sqrt{(s+a^2)(s+b^2)(s+c^2)} \qquad (s = \xi, \eta, \zeta) \tag{36}$$

To find the potential of a conducting ellipsoid of semiprincipal axes a, b and c imbedded in a dielectric, note that a family of ellipsoids with a parameter ξ confocal with the desired ellipsoid can be constructed. In such an ellipsoidal coordinate system, with $\xi = 0$ representing the surface of the conductor, the potential ψ must be independent of η and ζ. If a function depending only on ξ which satisfied Eq. (35) and goes to zero at infinity can be found, this is the potential. We immediately see this function must satisfy

$$\frac{\partial}{\partial \xi}\left(R_\xi \frac{\partial \psi}{\partial \xi}\right) = 0 \quad , \quad R_\xi = \sqrt{(\xi + a^2)(\xi + b^2)(\xi + c^2)} \quad . \tag{37}$$

On integrating and choosing the constant to satisfy the condition $V \to \frac{e}{4\pi\epsilon\, r}$ as $r \to \infty$, a potential at any point ξ is found in the well-known form

$$V(\xi) = \frac{e}{8\pi\epsilon} \int_\xi^\infty \left[(\xi + a^2)(\xi + b^2)(\xi + c^2)\right]^{-\frac{1}{2}} d\xi \quad . \tag{38}$$

From this equation it can be seen that the capacity is given by

$$C = 8\pi\epsilon \left\{ \int_0^\infty \left[(\xi + a^2)(\xi + b^2)(\xi + c^2)\right]^{-\frac{1}{2}} d\xi \right\}^{-1} \tag{39}$$

and the surface charge density, which is given by minus ϵ times the normal gradient of the potential at the surface ($\xi = 0$) is given by

$$\sigma = \frac{e}{4\pi abc} \left[\frac{x^2}{a^4} + \frac{y^2}{b^4} + \frac{z^2}{c^4}\right]^{-\frac{1}{2}} \quad . \tag{40}$$

If $a > b = c$ the spheroid is prolate, and the capacitance is given by

$$c = 8\pi\epsilon \left\{ \int_0^\infty \frac{d\xi}{\sqrt{\xi + a^2}\,(\xi + b^2)} \right\}^{-1} = 8\pi\epsilon a \, \frac{\ell}{\ln\left(\frac{1+\ell}{1-\ell}\right)} \quad . \tag{41}$$

Comparing this equation to Eq. (16), and using Eq. (12), we derive Eq. (33). Here ℓ is the eccentricity of the spheroid $\left(\ell = \sqrt{1 - (b/a)^2}\right)$.

A prolate spheroid is spherical when $\ell = 0$, rod-shaped when $\ell = 1$, and cigar shaped in between.

When $a = b > c$, the spheroid is oblate, and

$$c = 8\pi\epsilon \left\{ \int_0^\infty \frac{d\xi}{(\xi + a^2)\sqrt{\xi + c^2}} \right\}^{-1} = 4\pi\epsilon \sqrt{a^2 - c^2} \left[\tan^{-1}\sqrt{\frac{a^2 - c^2}{c^2}}\right]^{-1} \tag{42}$$

18

This geometry can approximate an island as shown in Fig. 4. When C = 0 this becomes a circular disk, and the capacitance becomes $8\epsilon a$, so that the activation energy is $\pi/2$ times as large as expressed by Eq. (17).

When the island, which we could assume for the moment to be a circular disk, is on a substrate, the capacitance will be further modified to

$$C = 4a(\epsilon_\ell + \epsilon_0) \tag{43}$$

where ϵ_ℓ is the low frequency dielectric constant of the substrate, and ϵ_0 is the permittivity of vacuum. In this case the electrostatic energy is given by

$$E_c' = \frac{e^2}{8a(\epsilon_\ell + \epsilon_0)} \quad . \tag{44}$$

The surface charge is distributed between the top and the bottom surfaces in the ratio of ϵ_0/ϵ_ℓ. Applying this fact to Eq. (40), which was originally derived for immersion in a uniform dielectric, it can be seen that the surface charge on the top of the disk will be given by

$$\sigma = \frac{\epsilon_0}{\epsilon_\ell + \epsilon_0} \frac{e}{4\pi ab} \left[1 - \frac{x^2}{a^2} - \frac{y^2}{b^2}\right]^{-\frac{1}{2}} \quad . \tag{45}$$

Thus, when we are on the circumference of the disk, σ becomes infinite, and from Eq. (21) the outward stress on the metal from this charge also becomes infinite. From Eq. (41) it can be seen that if V is the volume of the ellipsoid, and if c is small but non-zero, the surface charge density at the semi-axes $(a,o,o,)$ and (o,b,o) is

$$\sigma(a,o,o) = \frac{\epsilon_0}{\epsilon_0 + \epsilon_\ell} \frac{3e}{V} a \quad , \qquad \sigma(o,b,o) = \frac{\epsilon_0}{\epsilon_0 + \epsilon_\ell} \frac{3e}{V} b \tag{46}$$

so that the charge density, and, hence, the surface stress, is largest at the end of the principal semi-axis.

From the preceding discussion it may be supposed that an ellipsoid would tend to elongate when a charge resides on it. To check this assumption, the rate of charge of electrostatic energy with the semi-axes would have to be calculated, since if the volume remains constant, extension along the major semi-axis would have to be accompanied by contraction along one of the other semi-axes which might tend to cancel out any lowering of the electrostatic energy. Ignoring for the time being any difference in the dielectric constant between the substrate

19

and the vacuum, we can consider for simplicity an ellipsoidal disk whose capacitance can be calculated from Eq. (39) with c = 0. The volume can be taken to be proportional to πab, so that the capacitance can be written as

$$C_e = 8\pi c \left\{ \int_0^\infty \left[(a^2+\xi)\left(\frac{V^2}{\pi^2 a^2} + \xi\right)\xi \right]^{-\frac{1}{2}} d\xi \right\}^{-1} \qquad (47)$$

Taking the derivative of C_e with respect to a,

$$\frac{dC_e}{da} = \frac{C_e^2}{8\pi c} \left\{ \int_0^\infty \frac{1}{2} \left[(a^2+\xi)\left(\frac{V^2}{\pi^2 a^2} + \xi\right)\xi \right]^{-3/2} \xi \right.$$

$$\left. \left[2a\left(\frac{V^2}{\pi^2 a^2} + \xi\right) - (a^2+\xi)\frac{2V^2}{\pi^2 a^3} \right] d\xi \right\} \ . \qquad (48)$$

Every term in the integrand is positive with the possible exception of the last term, and dC_e/da will be positive or negative depending on the sign of

$$2a\left(\frac{V^2}{\pi^2 a^2} + \xi\right) - (a^2+\xi)\frac{2V^2}{\pi^2 a^3} = 2\xi\left(a - \frac{V^2}{\pi^2 a^3}\right) = 2\xi\left(a - \frac{b^2}{a}\right) \ . \qquad (49)$$

Since capacitance is inversely proportional to electrostatic energy, as C increases, E_c decreases, and visa versa. When a = b, i.e., the disk is circular, the electrostatic energy is at a maximum, and if one of the semi-axes is increased, the electrostatic energy will be decreased. According to Eq. (46) there will be an increased stress on the larger semi-axis tending to further extend it, and from Eqs. (48) and (49) this extension will lead to a further lowering of the electrostatic energy. Thus, a charged island will tend to distend until the increased surface energy due to the increased surface area equals the decreased electrostatic energy.

If a is the principal semi-axis, then from Eqs. (46) and (21), the stress at (a,o,o) is

$$P = \frac{9}{2} \frac{\epsilon_0 e^2 a^2}{(c_\ell+\epsilon_0)^2 V^2} = \frac{2.96}{b^2 c^2} \times 10^{10} \frac{kg}{m^2} \ , \qquad (50)$$

where c_ℓ has been taken to be $3.9\,\epsilon_0$ and b and c are expressed in Å. Since the yield point for pure cold-drawn gold is 2.5×10^7 kg/m^2, it is possible to exceed this value for small cross-section gold islands.

20

Even when the yield point is not exceeded, surface diffusion could cause distortion of the island. This force may be one cause for the filamentary nature of thicker films, since thicker films will have a larger percentage of the islands charged because of the low activation energy.

By the use of spherical harmonics, it can be shown[35] that a circular disk in a constant field $\mathcal{E}$ in the plane of the disk has a potential in oblate spheroidal coordinates of

$$V = - \frac{2}{\pi} \mathcal{E} \rho \left(\tan^{-1} \xi + \frac{\xi}{1+\xi^2} \right) \cos \varphi \quad , \qquad (51)$$

where ρ is the radial distance from the x axis (the minor semi-axis), ξ is the ellipsoidal coordinate and φ is the angle between the $\mathcal{E}$ field and ρ. If the applied field is in the y direction, and a is the radius of the disk, then in the y direction in the plane of the disk we have

$$V_y = - \frac{2}{\pi} \mathcal{E}_y \left\{ \tan^{-1} \left[\left(\frac{y}{a} \right)^2 - 1 \right]^{1/2} + \frac{[(y/a)^2-1]^{1/2}}{(y/a)^2} \right\} \quad . \qquad (52)$$

The potential in the y direction with x = 0 due to a charge e on the disk is

$$V_y = - \frac{e}{4\pi\epsilon_\ell a} \tan^{-1} \left[\left(\frac{y}{a} \right)^2 - 1 \right]^{-1/2} \quad . \qquad (53)$$

Since the total potential is the sum of these two expressions, the maximum in the potential barrier is determined by

$$\frac{dV_{TOT}}{dy} = - \frac{2}{\pi} \mathcal{E} \left\{ \tan^{-1} [(y/a)^2 - 1]^{1/2} + 2 [(y/a)^2 - 1]^{-1/2} - (y/a)^2 [(y/a)^2 - 1]^{1/2} \right.$$

$$\left. + \frac{e}{4\pi\epsilon a^2} [(y/a)^2 - 1][(y/a)((y/a)^2 - 1)^{3/2}]^{-1} = 0 \quad . \qquad (54)$$

If $y_{max} \gg a$, then this can be approximated by

$$\frac{dV_{TOT}}{dy} \approx - a\mathcal{E} + \frac{e}{4\pi\epsilon y^2} = 0 \qquad (55)$$

The maximum in the potential is accordingly

$$y_{max} \sim \sqrt{\frac{e}{4\pi \epsilon \mathcal{E}}} \qquad (56)$$

and the lowering of the potential is

$$\Delta V \approx -2 \sqrt{\frac{e \mathcal{E}}{4\pi \epsilon}} \qquad (57)$$

This change of potential is of the same form as that found for a spherical conductor as given by Eq. (4). Such a result is to be expected, since at distances considered large compared with the dimensions of a charged conductor, the shape of the conductor becomes immaterial. It can be shown, however, that for physically realizable fields, the assumptions involved in going from Eq. (54) to Eq. (55) become questionable. For example, if $\mathcal{E}$ is 10^5 V/cm the maximum field used in the experiments, the value of y_{max} given by Eq. (56) is 40 Å. For almost any reasonable value of a, all of Eq. (54) should be used, and V will not have the $\mathcal{E}^{1/2}$ dependence shown in Eq. (57). As shown in Fig. 21, the experimental results show that the current has a $\mathcal{E}^{1/2}$ dependence in this region. It should also be noted that for low fields, when $y_{max} \gg a$, there are intervening islands between y_{max} and the charged island, which, in a manner analogous to the previous development for the line of sphe s, will lead to the potential increasing significantly slower than would be expected from a straight-forward application of Eq. (53). This should also affect the $\mathcal{E}^{1/2}$ dependence at lower fields, contrary to experiment.

Although not mentioned previously, rigorous calculation of the field lowering around spherical conductors should include the perturbation of the field in the neighborhood of the sphere, so that terms analogous to those appearing in Eq. (54) should also appear in the calculations leading to Eq. (4). This problem should appear in all the earlier, more simplified calculations.

To determine the interaction of charged disks, we can resort to the results of E. W. Hobson[16] who determined the effect of an external charge on a grounded circular disk. Hobson found the Green's function in peripolar coordinates, and used the Green's function to determine the potential through all space caused by the interaction of the charge and the conducting disk. He found when the charge q is in the plane of the film at a point P, and we are interested in the induced surface charge density, σ, at a point Q, that σ is given by

$$\sigma(Q) = -\frac{q}{2\pi^2} \frac{1}{\overline{PQ}^2} \sqrt{\frac{\overline{CP}^2 - a^2}{a^2 - \overline{CQ}^2}} \quad , \qquad (58)$$

In this expression a is the radius of the disk, $\overline{PQ}$ and $\overline{CP}$ are the distances from the external charge to the point where σ is determined and the center of the disk, respectively, and $\overline{CQ}$ is the distance from the point where p is measured to the center of the disk. The total induced charge on the disk is $-q/\pi$ times the angle between the tangents from the charge to the circular boundary of the disk. Therefore the total induced charge is

$$Q = -\frac{2q}{\pi} \sin^{-1} \frac{a}{\overline{PC}} \quad . \tag{59}$$

It can be seen, therefore, that if we have an insulated circular disk, the induced surface density for $\epsilon_0 = \epsilon_\ell$ will be given from Eqs. (45), (58), and (59) as

$$\sigma(Q) = \frac{1}{2\pi^2} \frac{1}{\sqrt{a^2-\overline{CQ}^2}} \left\{ q \left[\frac{\sin^{-1}\frac{a}{\overline{PC}}}{2a} \frac{\sqrt{\overline{CP}^2-a^2}}{\overline{PQ}^2} \right] + \frac{\pi C}{4a} \right\} \tag{60}$$

where C is the total charge of the disk. Also, if an uncharged disk is in the presence of a charge q, such as a charge on an adjacent disk, and if $\overline{CQ} = a$ so that we are on the rim of the disk, then σ becomes infinite. Hence, the outward stress on the disk edge becomes infinite, leading to extension of the disk. In practice, this force would not become infinite but would be inversely proportional to the radius of curvature of the edge of the cap, still leading to very large forces.

The case of two charged disks cannot be treated as easily as was done for two charged spheres because of the absence of a single discrete image charge. A complicated approach would be possible by having an infinite multiple integration of the surface charge densities as expressed by Eq. (58), but the complexity does not seem to yield any further information.

It may be noted that peri-polar coordinates have, as one set of orthogonal coordinates, a system of spherical bowls with the fundamental circle as a common rim, including the disk discussed above. Therefore, the cap form of island usually considered the best approximation to the real case can be effectively approached by the use of this coordinate system.

SECTION IV

CONDUCTION PROCESSES

When considering the flow of current through discontinuous films, it is generally considered that the charge carriers are supplied by and travel between the metallic "islands" at the surface of the dielectric. Unless charges are simultaneously transported along a line of islands, this conduction process requires the alternate charging and discharging of individual islands. There are two interrelated problems, each with its own distinct considerations; the creation of a charged island, and the transport of that charge from one island to the next. The latter will be discussed first.

The motion of charges from one island to the next is impeded by an inter-island barrier. The barrier may be associated with the work function,[4] the contact potential,[27] or the discrete island size.[22] The charge must either tunnel through the barrier or surmount it by thermal excitation. The theoretical choice between these two modes of transport traditionally has revolved about the form of the potential barrier, tunneling being favored for high, thin barriers, and thermal activation for low, thick barriers. In the following discussion it will be shown that neither of these possibilities can be considered without considerable modification.

Considering tunneling first, the expected velocity along the direction of the electric field can be written as

$$< v_x > \alpha \int_0^{E_{max}} D(E_x) \left\{ \int_0^{\infty} [f(E) - f(E-eV)] \, dE_y \right\} dE_x \qquad (61)$$

where $f(E)$ is the Fermi function, E, E_x, and E_y are the total, x-component kinetic energy, respectively, and $D(E_x)$ is the WBK tunneling probability.

$$D(E_x) = \exp \left\{ -\frac{4\pi}{h} \int [2m(v_x - E_x)]^{1/2} dx \right\} \qquad . \qquad (62)$$

One very weak point in these two equations is that they were derived in one dimension. Logically one should not proceed from the one-dimensional case to a three-dimensional case with complex geometries without justification. If the problem is radially symmetric, for example, separation of variables gives Schroedinger's equation for the radial function R which can be reduced to a one-dimensional form by the substitution $R = u/r$. In the case where there is no radial symmetry, however, this cannot be done without altering the form of the radial potential. At

25   **Preceding page blank**

the very least, if Eq. (62) is evaluated by integrating over the values of x corresponding to the inter-island separation, Eq. (62) is valid for only the infinitesimally small area on each island closest to the opposite island. The entire problem could be treated on a more exact theoretical level by the use of Feynman path integrals, but this has not been done according to the literature searched for this work.

Equation (1) has been considered by Simmons[37] who solved it in the form

$$< v > \propto \frac{1}{B^2} \frac{\pi BkT}{\sin(\pi BkT)} \exp[-A \bar{\varphi}^{\frac{1}{2}}](1 - \exp[-BeV])$$ (63)

where $\bar{\varphi}$ is the average barrier height above the Fermi level,

$$A = \frac{4\pi \Delta s}{h} (2m)^{\frac{1}{2}}$$ (64)

and

$$B = \frac{A}{(2\bar{\varphi})^{\frac{1}{2}}} ;$$ (65)

Δs is the distance between the points where the potential barrier meets the Fermi levels. Equation (63) illustrates the salient features of tunneling conduction: (1) a weak temperature dependence, (2) a strong dependence on the barrier width, Δs, and (3) strong dependence on field. Usually for any barrier width greater than about 50 Å, the tunneling probability becomes so small as to be negligible. This makes especially pertinent the problem of what fraction of the area of an island can contribute to conduction due to its proximity to the next island.

As seen in Eq. (63), the exponential dependence of the resistance on reciprocal absolute temperature as illustrated in Fig. 14 is not characteristic of tunneling. The theories incorporating tunneling have approached this problem from several directions.

Herman and Rhodin[11] considered the discrete levels present in the energy spectrum of the metal due to the microscopic island size. As the energy is no longer quasi-continuous, for an electron to tunnel from one island to an unoccupied level in an identical adjacent island it is necessary for the electron to first be excited to an activated state. Tunneling, therefore, requires an amount of thermal energy equal to the energy separation between allowed states in the metal. Since the energy separation increases with decreasing island size, thinner films should, in general, require a larger amount of thermal energy. This thermal activation is held by Herman and Rhodin to be the cause for the exponential reciprocal temperature dependence. While the effect described by

Herman and Rhodin undoubtedly exists, the high activation energies experimentally observed are not easily accounted for by this theory, and this theory does not include the electrostatic energy of the charged islands.

Neugebauer[19],[20] and Hill[27] have considered thermally-assisted tunneling whereby, in order to create a charged island initially, an electron must be first thermally excited to an energy equal to the electrostatic energy. A simple treatment of this problem, such as the derivation of Eq. (4), shows that an applied field causes a lowering of the electrostatic energy requirement by a factor proportional to the square root of the electric field. This type of voltage dependence is contingent on the potential being inversely proportional to the distance from the charge source, and the islands being closely spaced with respect to the distance to the potential maximum. The latter requirement arises from the square root of electric field dependence being due to the movement of the barrier maximum with changes in the applied field. When a charge carrier tunnels over a fixed distance in an electric field the energy gained from the electric field is directly proportional to the applied voltage irregardless of the shape of the potential between the islands. If the potential maximum appears between two islands, the barrier lowering should be proportional to the applied voltage, rather than the square root of the voltage as in Fig. 21. The distance to the barrier maximum can be calculated from Eq. (3). Using 3.78 for the relative dielectric constant of fused quartz, the distance to the barrier maximum is $0.67 \times 10^{-4} \; \mathcal{E}^{-\frac{1}{2}}$ (V-cm)$^{\frac{1}{2}}$. This expression would result in the distance to the barrier maximum for the largest voltage in Fig. 21 to be on the order of 20 Å. From Fig. 21 the deviation from the square-root-of-applied-voltage-dependent behavior at this small distance to the barrier maximum has not been observed. At lower voltages, a deviation in the potential varying as the inverse distance from the charge source would be expected because of the shielding effect of intervening islands. This should also lead to a deviation from the square root of voltage dependence which has not been observed.

There is a more fundamental objection to carrier transport by tunneling, however, that has its origin in the application of macroscopic equations for the tunneling of an electron between two macroscopic plates separated by a dielectric. In the macroscopic system, considering the small value of r/σ in the metal, the loss of an electron will not affect the potentials and charge distribution significantly. This same assumption will have to be examined in the case of the microscopic dimensions encountered in a discontinuous film.

Consider temporarily the simple model of a metal island completely immersed in a dielectric. When that island is charged by a single electronic charge which has resided on the island for a long period of time, the charged island represents a certain electrostatic energy which has been calculated previously for various possible geometries. In all cases, that electrostatic energy is inversely proportional to the low

27

frequency dielectric constant ϵ_ℓ. The fact that $\epsilon_\ell > \epsilon_0$, i.e., the electrostatic energy is lower than would be seen if an identical island were identically charged in a vacuum, can be attributed to the transfer of electrostatic energy from the charge to the dielectric polarization of the medium. The amount of energy stored in the dielectric at low frequencies is

$$E^* = E_c \epsilon_\ell \left(\frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_\ell} \right) = E_c \frac{\epsilon_\ell}{\epsilon^*} \tag{66}$$

where ϵ_∞ is the high-frequency dielectric constant and ϵ_ℓ is the electrostatic energy calculated previously. ϵ_ℓ in a solid is usually dominated by the electronic polarizability and longitudinal phonon contributions. If a charge instantaneously appears on a disk-shaped conductor, for example, the electrostatic energy immediately upon the appearance of the charge would be given by

$$E_c = \frac{e^2}{16\epsilon_\infty a} \tag{67}$$

After the dielectric relaxes the electrostatic energy is

$$E_{co} = \frac{e^2}{16\epsilon_\ell a} \quad . \tag{68}$$

Since no additional energy is supplied to the system during the relaxation the difference between these two energies must reside in the change of the polarization of the dielectric.

The time that an electron takes to tunnel across a potential barrier has been investigated by T. E. Hartman[38] who found it to be on the order of 10^{-16} seconds. Therefore, if an electron tunnels from an island with electrostatic energy represented by Eq. (68) to a distant island, on arrival, the conditions necessary for the derivation of Eq. (67) will be satisfied. An additional amount of energy given by the difference between Eq. (67) and Eq. (68) will remain in the dielectric around the original position of the electron for an instant after the tunneling. While the electrostatic energy of the system was given by Eq. (68) before tunneling, immediate after tunneling the electrostatic energy of the system will be

$$E_c = \frac{e^2}{16\epsilon_\infty a} + \left[\frac{e^2}{16\epsilon_\infty a} - \frac{e^2}{16\epsilon_\ell a} \right] = \frac{e^2}{16a} \left[\frac{2}{\epsilon_\infty} - \frac{1}{\epsilon_\ell} \right] \quad . \tag{69}$$

28

An amount equal to

$$\delta E = \frac{e^2}{16a}\left[\frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_\ell}\right] \tag{70}$$

must, therefore, be added to the system thermally before tunneling in order that the energy immediately after the tunneling be equal to the energy in the system before tunneling. In the case of fused silica, the low frequency dielectric constant is 3.78 ϵ_0, while at optical frequencies the index of refraction is 1.406. From these numbers, Eq. (70) can be written as

$$\delta E = .241 \frac{e^2}{16a\epsilon_0} = .911 \, E_{co} \quad , \tag{71}$$

where E_{co} is the initial energy in the system given by Eq. (68). For other dielectrics which have been used in the literature, the difference between the high and low frequency constant is even more pronounced, resulting in an even larger ratio of $\delta E/E_{co}$.

If the tunneling were through the vacuum between the islands rather than through the dielectric, some of the above objections would be apparently overcome. The potential barrier for such tunneling would be the work function of the metal. Tunneling through the substrate would have a barrier due to the contact potential between the metal and the dielectric, which is normally considerably less than the work function. In addition, tunneling through the vacuum would not basically alter Eq. (70) since most of the electrostatic energy before tunneling would still reside in the dielectric substrate.

Equation (70) was derived assuming infinite separation of the charges after tunneling, which is obviously a poor approximation. Tunneling a finite distance will reduce δE but will not change any fundamental arguments. The effect of tunneling between islands separated by a distance, D, can be seen best by using spherical islands and using the approximation that the sudden departure of a charge, e, from an island leaves a polarization charge

$$Q_{pol} = -e\left[1 - \frac{\epsilon_\infty}{\epsilon_\ell}\right] \tag{72}$$

The total energy of the system immediately after tunneling can then be found from Eqs. (28), (17), and (29) to be

29

$$E_{c_{TOT}} = \frac{e^2}{8\pi\epsilon_o a} + \frac{e^2\left[1 - \frac{\epsilon_\infty}{\epsilon_o}\right]^2}{8\pi\epsilon_\infty a}$$

$$- \frac{e^2\left[1 - \frac{\epsilon_\infty}{\epsilon_\ell}\right]}{8\pi\epsilon_\infty a} \frac{\frac{2}{\sinh\beta}\sum\limits_{n=1}^{\infty} \text{csch } 2n\beta}{\left[\sum\limits_{n=1}^{\infty} \text{csch }(2n-1)\beta\right]^2 - \left[\sum\limits_{n=1}^{\infty} \text{csch }2n\beta\right]^2} \qquad (73)$$

where β is given by Eq. (27). This can be approximated for larger D/a by

$$E_{c_{TOT}} \approx \frac{e^2}{8\pi\epsilon_\infty a}\left\{\left[1 + \left(1 - \frac{\epsilon_\infty}{\epsilon_\ell}\right)^2\right] - \left(1 - \frac{\epsilon_\infty}{\epsilon_\ell}\right)\frac{1}{D}\right\} . \qquad (74)$$

It may be noted that this equation is larger for $D \to \infty$ than would be expected from Eq. (70), because the preliminary calculation neglected the extra energy required for the electron to escape from the attractive force of the polarization field.

This analysis is the analogue of the Frank-Condon[39] principle which accounts for the difference between the optical and thermal dissociation energies by assuming that in the former case the coordinates of the lattice particles do not change during the transition, and also that the relative velocities between the lattice particles remains unchanged.

Another situation where the dielectric relaxation effect can be seen[40] is in the tunneling characteristics of polar semiconductors, as shown in Fig. 5. At low temperatures, the conductance of the tunneling current through these materials shows a minimum at zero applied volts that has been attributed to the inability of the polarization field to follow the tunneling electron, and hence, the necessity of supplying the polarization or deformation field energy from the applied electric field.

It has, therefore, been shown that tunneling by itself is unacceptable as a means of charge transport, both because it does not explain the thermal activation energy and $v^{\frac{1}{2}}$ dependence of the thermal activation energy experimentally observed, and because tunneling introduces additional energy requirements which would make tunneling less energetically favorable vis-a-vis other transport mechanisms to be proposed.

One modification of tunneling transport is a model proposed by Hill[41] whereby an electron does not tunnel directly from one island the next, but rather makes the transition by means of intermediate s ps at any number of traps, or localized states between the islands. This model has an advantage in the present discussion since the tunneling
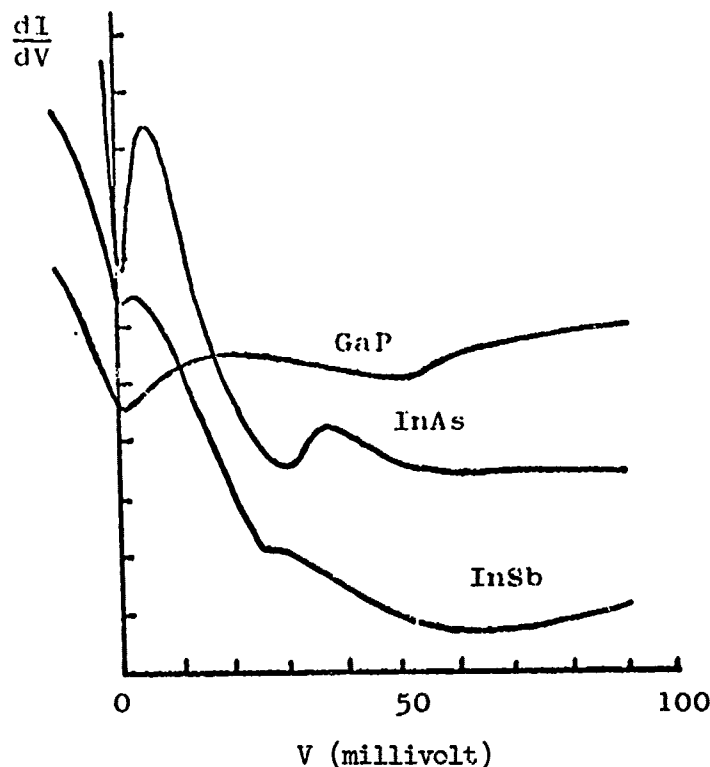
$\dfrac{dI}{dV}$

GaP

InAs

InSb

0   50   100

V (millivolt)

Fig. 5 - Tunneling Through Several 3-5 Material Junctions[40]

distance is small, and hence the additional energy required, because the polarization field cannot follow the electron, is reduced since the electron does not have to tunnel all the way out of the polarized dielectric. Any trap which is positively charged when empty (such as sodium ions in the borosicicate glasses Hill considered) would enhance tunneling probability across the inter-island gaps because the depression in the potential barrier that the electron would encounter would lower the average barrier height $\bar{\varphi}$ appearing in Eq. (63), Hill used this model to correlate the effect of ion migration in the substrate to the film conductivity under the influence of an external bias. This modified model, however, shares with direct tunneling the inability to account for the $\mathcal{E}^{\frac{1}{2}}$ dependence of the thermal activation energy.

Thermionic emission, whereby an electron gains enough thermal energy while inside an island to surmount the potential barrier between islands has been proposed by Minn[25] and van Steensel.[26] In the analysis by van Steensel, the current transport between two islands is likened to that between two planar electrodes, in which the potential can be established by the method of images between the electrodes, with the boundary condition that the potential go to zero at an infinite distance from the electrodes. The infinite series thus established can be solved for

31

an electron at the midpoint between two electrodes to give a potential
lowering of

$$
V = \frac{e^.}{2\pi^. s} \sum_{n=1}^{\infty} \frac{(-1)^n}{n} = -\frac{e^.}{2\pi^. s} \gamma 2 = -\frac{20}{Ks(\text{in }\text{Å})} \text{ eV} , \qquad (75)
$$

where s is the separation between the electrodes and k is the relative
dielectric constant. While this potential can be considered to be the
integral of the electric field, most of the potential contribution is
due to the region adjacent to the substrate. In this region, the elec-
tron with thermal energy high enough to surmount the barrier will be
traveling too fast for the dielectric polarization field to follow it,
and, hence k must be the high-frequency relative dielectric constant.

van Steensel considers the potential barrier to be the difference
between the metal work function and the electron affinity of the dielec-
tric, reduced by Eq. (75). In the presence of a small electric field,
ϵ, the barrier height will be reduced by an amount ϵ s/2, which
van Steensel concedes to be in contradiction with the observed reduc-
tion in barrier height proportional to $\epsilon^{\frac{1}{2}}$.

R. M. Hill[27] has also considered the shape of the potential barrier
considering the image force from a spherical island. Using a computer
solution to the multiple image problem, he concludes there is little
difference between the barrier between two spheres and between two
planar electrodes. This investigator has found the opposite to be the
case. When an electron is outside an uncharged sphere, the force of
attraction to the sphere is given by Eq. A-1 with Q = 0. Therefore,
the amount that the potential is lowered from that at an infinite dis-
tance away when the electron is a distance R from the center of the
sphere is

$$
\Delta \varphi = \int_{\infty}^{R} \frac{a^. e^.}{4\pi\epsilon} \left[\frac{2x^.-a^.}{x^3(x^.-a^.)^2} \right] dx = \frac{a^3 e^.}{8\pi\epsilon} \frac{1}{R^2(R^2-a^2)} . \qquad (76)
$$

Therefore, neglecting multiple imaging, the difference between the bar-
rier height at an infinite distance and the barrier height at the center
of the gap between two spheres separated by a distance D can be found
by calculating the reduction due to each of the spheres. This calcula-
tion gives

$$
\Delta \varphi_{GAP} = \frac{4a^3 e^2}{\pi\epsilon D^2(D^2-4a^2)} . \qquad (77)
$$

32

In practical units, with a and D expressed in angstroms,

$$\Delta \varphi_{GAP}(\text{in eV}) = \frac{230a^3}{KD^2(D^2-4a^2)} \tag{78}$$

If the surface of two islands of 50 Å radius are separated by 50 Å, Eq. (78) gives $\Delta \varphi$ as $\frac{0.102}{K}$ eV, while Eq. (75) predicts $\frac{0.4}{K}$ eV for planar electrodes with the same separation. If the separation of the island surfaces increases to 100 Å, Eq. (78) predicts a barrier lowering of $\frac{0.024}{K}$ eV, while Eq. (75) predicts $\frac{0.2}{K}$ eV barrier lowering for similarily spaced planar electrodes. It should be noticed that the neglect of multiple imaging in the derivation of Eq. (77) would result in a lower $\Delta \varphi$ than would actually exist, although the short range of Eq. A-1 would make this error quite small. Hill did not give the details of his calculation.

As was mentioned previously, in Eqs. (75), (77), and (78) K is the high frequency dielectric constant since, in the vicinity of the conductor, the electron moves too fast for the polarization field to follow. The consequence of this is that all the comments concerned with the extra required activation energy for tunneling apply equally for thermionic emission.

It can be seen from the preceding discussion that image effects from an uncharged island are very short-ranged compared to the same effect from a planar surface. The basic reason for this is a charge outside an infinite surface will attract an opposite charge to the surface of the conductor. A compensating charge in the conductor is pushed an infinite distance away to retain charge neutrality in the conductor. Because of this infinite separation, the compensating charge does not enter into the calculations. In a small sphere, on the other hand, this compensating charge must be placed at the center of the sphere and thus cancels out most of the effect of the induced charge at moderate distances from the sphere.

If the potential barrier is not greatly lowered by the image force, then the energy that an electron must be excited to in thermionic emission is, as reported earlier, the difference between the metal work function and the dielectric electron affinity. The work of Deal, Snow, Mead[42] has shown that both the high frequency photo-response barrier and the low-frequency capacitance-voltage curve barrier for an electron leaving gold and entering the conduction band of thermally oxidized silicon is 4.1 eV. This value is in conflict with the low activation energies found experimentally for discontinuous gold films on many types of glass substrates. The same problem exists for the other metals. Some method of drastically lowering the potential barrier is required, and image force lowering does not suffice.

Since there are fundamental problems associated with the assumption of either tunneling or thermal emission, a new model will be proposed to overcome these problems. It will be assumed that it is not necessary to raise an electron to the conduction band of the insulator substrate. A sufficient density of trap states around the Fermi level of the insulator will allow charges to exist between islands. In this case, slow carrier transport would then be possible, and the only barrier to the transport would be the charge's electrostatic force of attraction to a near-by neutral island. To justify this method of charge transfer, the energy level structure in the substrate will be described in the following section.

CHAPTER V

ISLAND-SUBSTRATE ENERGY LEVELS

The energy structure of the electrons in the system under consideration will first be examined. This is a very complicated problem. The substrate, quartz, is normally amorphous, with the attendant distortions of the energy band especially apparent in the blurring of the band edges caused by an admixture of traveling and decaying (localized) wave function solutions to Schroedinger's equation. In addition, there is generally a large concentration of trapping centers in normal glass which reduces the conductivity. To complete the picture, the part of the glass under consideration is the surface, with its attendant restructuring, electronic surface states and contamination. In the metal we have the problems inherent in determining the normal electronic states, together with the large surface-to-volume ratio increasing the importance of surface states, which renders many of the normal techniques inappropriate. Additionally, the finite number of atoms involved requires, in many cases, the consideration of the individual, discrete states rather than the continuum of states better approximated by macroscopic samples. Finally, the interaction between the substrate and the metal islands must be considered. Again the finite size of the islands produces interesting effects. All these complications will require approximation and generalization.

Considering the metal first, there are several ways to approach the problem. Ideally, a many-body calculation should be used which sets up a Hamiltonian representing all the particles in the system and all the interactions among the particles. In general, this calculation turns out to be impossible because of the number of variables involved. The problem is usually approached by reducing it to a one particle calculation using the symmetry involved in an infinite crystal. A crystal is broken up into unit cells, and electron states are calculated inside the unit cell with symmetrical boundary conditions. A simplified one-dimensional calculation along this line is the Kronig-Penny calculation.[43] Normally these one-electron states are made antisymmetrical so as to obey the Pauli exclusion principle. This approach leads to the Hartree-Fock equations, where all the electrons have their energy adjusted by interaction with the field of an exchange or Fermi hole which accompanies the electron and has an electronic charge when looking at an occupied state and zero charge when looking at an excited state. Since they are not being repelled by the Fermi hole, the energy of the occupied states tends to be lowered by an amount known as the exchange energy.

It was found,[44] however, that the best Hartree-Fock calculations gave binding energies which were not large enough. The trouble was attributed to the fact that there was no simple way, using one electron states, to completely account for the coulombic repulsion between

electrons. In real life the minimum energy state of an electron is dependent upon the states of all the other electrons, so that the electron states are correlated and cannot be derived by considering an average interaction as is done in the Hartree-Fock calculations. For this reason a further energy correction, called the correlation energy, was introduced. As Slater[45] states, "the proper definition of the correlation energy is the difference between the Hartree-Fock energy of a system, and the true experimental energy." While there has been agreement as to what the correlation energy is, there has been less agreement in its mathematical expression, and no completely satisfactory closed form of the expression has been found. For the case of an electron outside the surface of a metal, it is the correlation energy that would give the image force on the electron since this is just another example of the presence of one state (the electron outside the metal) having a reciprocal effect on other states (the electrons pushed away from the surface of the metal) and modifying the waveforms which give the lowest total energy. The ground state energies and corrections in the limit of a dense electron gas can be calculated most elegantly by the use of Feynman diagrams.

We will leave this discussion for the time being, keeping in mind that there are adjustments to the simple one electron states, which tend to move these states as a group and possibly to spread them out or contract them. The simplest qualitative calculation of these one electron states is from the one-dimensional Kronig-Penney model.[43]

In order to investigate the effect of the limited physical dimension of an individual metal island on the energy levels of the electrons, let us construct a one-dimensional finite Kronig-Penney type model.

We want an electron wave function which will satisfy the one-dimensional Schroedinger equation

$$\frac{d^2}{dx^2} \psi + \frac{2m}{\hbar^2} (E - V) \psi = 0 \tag{79}$$

where V is the potential energy (as a function of x) and E is the energy of the electron. We can consider a potential given as follows (which represents a $2N + 1$ ion cores with potential $-(2\hbar^2 Z/m)^{1/2}$ separated by a distance aN).

$$V(x) = \begin{cases} 0 & x < -aN \quad x > aN \\ -(2\hbar^2 Z/m)^{1/2} \, \delta(x-an) & -aN \leqslant x \leqslant aN \quad n = 0, 1, 2 \cdots \pm N . \end{cases} \tag{80}$$

The reason for the strange normalization in V and the identity of Z will become apparent later.

First write Eq. (79) in the following form:

$$\left(\frac{d^2}{dx^2} + \frac{2mE}{\hbar^2}\right)\psi = \frac{2m}{\hbar^2} V\psi \tag{81}$$

A Green's function $(G(x,\xi)$ which satisfies the differential equation

$$\left(\frac{d^2}{dx^2} + \frac{2mE}{\hbar^2}\right)G(x,\xi) = \zeta(x-\xi) \tag{82}$$

is next defined, subject to the boundary condition that G represent only outgoing waves at infinity. G(x) can be easily found by taking the Fourier transform of $\mathscr{A}(x)$

$$G(x,\xi) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathscr{A}(k,\xi) e^{ikx} dk \quad . \tag{83}$$

Substituting this equation into Eq. (82) results in

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \left(-k^2 + \frac{2mE}{\hbar^2}\right) \mathscr{A}(k,\xi) e^{ikx} dk = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ik(x-\xi)} dk$$

$$\therefore \quad \mathscr{A}(k,\xi) = \frac{e^{-ik\xi}}{\dfrac{2mE}{\hbar^2} - k^2}$$

$$\therefore \quad G(x,\xi) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{ik(x-\xi)}}{\dfrac{2mE}{\hbar^2} - k^2} dk = \frac{i}{2\sqrt{\dfrac{2m}{\hbar^2}E}} e^{i(x-\xi)\sqrt{\frac{2mE}{\hbar^2}}} \tag{84}$$

In the integration in the last step of Eq. (84) it is necessary to keep in mind the boundary condition on G.

Applying Green's theorem to the functions $\psi(\xi)$ and $G(x,\xi)$ we have

$$\int_{-\infty}^{\infty} \left(\psi \frac{d^2 G}{d\xi^2} - G \frac{d^2\psi}{d\xi^2}\right) d\xi = 0 \quad . \tag{85}$$

37

$$\therefore \quad \psi(x) = -\frac{2m}{\hbar^2} \int_{-\infty}^{\infty} V(\xi) G(x,\xi) \psi(\xi) d\xi \qquad (86)$$

By substituting Eq. (80) for $V(\xi)$ in Eq. (86) we obtain

$$\psi(x) = -\frac{2m}{\hbar^2} \sum_{n=-N}^{N} -\left(\frac{2\hbar^2}{m} Z\right)^{1/2} \frac{1}{2\sqrt{\frac{2m}{\hbar^2} E}} e^{i|x-na|\sqrt{\frac{2mE}{\hbar^2}}} \psi(an)$$

$$\therefore \quad \psi(x) = \left(\frac{Z}{E}\right)^{1/2} \sum_{n=-N}^{N} i e^{i\left(\frac{2mE}{\hbar^2}\right)^{1/2}|x-na|} \psi(an) \quad . \qquad (87)$$

In general, this represents a set of $2N + 1$ equations in $2N + 1$ unknowns. The condition that the determinant of the $\psi(an)$ coefficients be zero gives the allowed energy values, e.g.,

$$\text{DET} \left\{ \sum_{n=-N}^{N} \left[\left(-\frac{Z}{E}\right)^{1/2} e^{i\left(\frac{2mE}{\hbar^2}\right)^{1/2}|m-n|a} - \delta_{mn} \right] \right\} = 0$$

$$-N \leq m,n \leq N \qquad (88)$$

If $N = 0$ we have an isolated ion core, and Eq. (88) corresponds to

$$\left[\left(-\frac{Z}{E}\right)^{1/2} - 1 \right] = 0$$

$$\therefore \quad E = -Z \qquad (89)$$

In this case we see that the energy of the electron is $-Z$. Hence Z must be the ionization potential of the individual metal atom by Koopman's theorem.[46] This is the justification for the original choice of normalization.

In the other direction consider what happens as $N \to \infty$. This is the case of an infinite crystal with periodic potential, hence by Bloch's theorem,[47] the solutions to Eq. (87) can be given by

$$\psi(am) = e^{ikam} \psi(0)$$

Therefore from Eq. (9)

$$\psi(0) = \left(-\frac{Z}{E}\right)^{1/2} \sum_{n=-N}^{N} e^{i\left(\frac{2mE}{\hbar^2}\right)^{1/2}|na|} e^{ikan} \psi(0)$$

$$\therefore 1 = \left(-\frac{Z}{E}\right)^{1/2} \sum_{n=-N}^{N} e^{i\left(\frac{2mE}{\hbar^2}\right)^{1/2}|na|} e^{ikan}$$

$$= \left(-\frac{Z}{E}\right)^{1/2}\left[\sum_{n=-\infty}^{0} e^{i\left[k-\left(\frac{2mE}{\hbar^2}\right)^{1/2}\right]an} \right.$$

$$\left. + \sum_{n=0}^{\infty} e^{i\left[k+\left(\frac{2mE}{\hbar^2}\right)^{1/2}\right]an} - 1\right]$$

$$\therefore 1 = \left(-\frac{Z}{E}\right)^{1/2}\left[\frac{1}{1-\exp\left\{(-ia)\left[k-\left(\frac{2mE}{\hbar^2}\right)^{1/2}\right]\right\}}\right.$$

$$\left. + \frac{1}{1-\exp\left\{ia\left[k+\left(\frac{2mE}{\hbar^2}\right)^{1/2}\right]\right\}} - 1\right]$$

$$1 = \left(-\frac{Z}{E}\right)^{1/2}\left[\frac{\sin(2mE/\hbar^2)^{1/2}a}{\cos ka - \cos\left(\frac{2mE}{\hbar^2}\right)^{1/2}a}\right]$$

$$\therefore \cos ka = \cos\left(\frac{2mE}{\hbar^2}\right)^{1/2}a + \left(-\frac{Z}{E}\right)^{1/2}\sin\left(\frac{2mE}{\hbar^2}\right)^{1/2}a \tag{90}$$

If $E < 0$, as has been considered,

$$\sin ix = i \sinh x$$

$$\cos ix = \cosh x$$

so that

$$\cos ka = \cosh\left(-\frac{2mE}{\hbar^2}\right)^{1/2}a - \left(-\frac{Z}{E}\right)^{1/2}\sinh\left(-\frac{2mE}{\hbar^2}\right)^{1/2}a \ . \tag{91}$$

This is the familiar Kronig-Penney solution. It can be easily solved for energy as a function of k. For example, if

$$\left(- \frac{2mE}{\hbar^2}\right)^{1/2} a > 1$$

so that

$$\sinh \left(- \frac{2mE}{\hbar^2}\right)^{1/2} a \approx \cosh \left(- \frac{2mE}{\hbar^2}\right)^{1/2} a \approx \frac{1}{2} \exp\left(- \frac{2mE}{\hbar^2}\right)^{1/2} a$$

then, from Eq. (91)

$$E \approx -Ze^{-\left(\frac{2mZ}{\hbar^2}\right)^{1/2} a} \cos ka \quad . \tag{92}$$

Familiarly from elementary crystal theory, this result shows that within the crystal an electron wave, as represented by the Bloch wave, is propagated freely along the array and is not localized on an individual atom as was the case for the isolated ion core. It can also be seen that there are electrons present with an energy

$$\Delta E = 4 Z e^{-\left(\frac{2mZ}{\hbar^2}\right)^{1/2} a} \tag{93}$$

above the ionization potential for an individual atom (provided the band is completely filled with electrons), which is not the case for the conduction band in a metal. These electrons can be more easily excited from the material when it coalesces into an array than when they are located on the individual, separated ion cores.

We have some confidence that Eq. (88) qualitatively represents conditions in a metal since for N = 0 we have shown that the results approximate individual metal atoms, and as N → ∞ this equation gives the familiar results of the Kronig-Penney model.

Equation (10) can be rewritten as

$$\mathrm{DET} \left\{ \sum_{n=-N}^{N} \left[e^{-|m-n|\kappa a} - \frac{\kappa}{b} \delta_{mn} \right] \right\} = 0 \tag{94}$$

where

$$\kappa = \left(- \frac{2mE}{\hbar^2}\right)^{1/2} \quad , \qquad b = \left(\frac{2mZ}{\hbar^2}\right)^{1/2} \quad . \tag{95}$$

Consider a numerical example with a = 2, b = 1.

With only one atom $1 - \frac{\kappa}{b} = 0$, X = 1 corresponding to E = - Z.

For two atoms the corresponding equation would be

$$\begin{vmatrix} 1 - \kappa & e^{-2\kappa} \\ e^{-2\kappa} & 1 - \kappa \end{vmatrix} = 0 \qquad 1 - \kappa = \pm\, e^{-2\kappa} \quad : \qquad (96)$$

$$\therefore \ \kappa = \begin{cases} 0.7968 \\ 1.09 \end{cases} \tag{97}$$

For the case of three atoms

$$\begin{vmatrix} 1 - \kappa & e^{-2\kappa} & e^{-4\kappa} \\ e^{-2\kappa} & 1 - \kappa & e^{-2\kappa} \\ e^{-4\kappa} & e^{-2\kappa} & 1 - \kappa \end{vmatrix} = 0 \tag{98}$$

$$\kappa = \begin{cases} 1.148 \\ 0.980 \\ 0.650 \end{cases} \tag{99}$$

For the case of N $\rightarrow \infty$, as discussed earlier

$$\kappa \approx 1 + 4e^{-2} \cos ka = 1 + 0.54 \cos ka \quad . \tag{100}$$

In this analysis we note that there are allowed energy states, and one or more values of k associated with each state. For one atom, the wavefunction ψ has an imaginary value of k, which means that the state is localized. As the number of electrons is increased the states become coupled, and the energy levels split. There is a possibility of an electron moving from one site to the next. In the limit, as the number of electrons becomes infinite, the electron can move from one site to another in a coherent manner, so that the eigen-value of k is real, representing a traveling wave. Each allowed value of k represents two states with opposite spin. In the case of monovalent atoms, such as the noble metals, there will be one electron per atom in the band, and since there are two states per atom, the states will be half-filled. This incomplete filling of bands, either by the methods just described or by the overlapping of different bands, is a necessary condition for metallic conduction. Various other ways of approximating the core potentials lead to the same general splitting of a discrete level into a multiplet of levels, one for each atom in the system. In general,

41

one would expect to find localized wave functions correspond' , to the surface atoms as in the analysis by Tamm[50] or Shockley,[51] whe. ' one state for each surface atom is detached to form localized states in the forbidden gap.

When the band structure is used to determine work functions or surface energies, or in any case where the absolute energy levels must be known, there is a further complication, the formation of an electrical dipole layer at the surface because of overlap of the electron density beyond the more rigid ion cores. J. R. Smith[48] has a very interesting treatment of this problem which starts, as we have discussed, by calculating a total energy as the sum of an interaction of the electrons without an outside potential (as from the ion cores) and with each other, together with the kinetic, exchange and correlation energies, all expressed in terms of the particle density function. He further expands the exchange and correlation energies as a gradient expansion in the particle density to account for density fluctuations at the surface. Since the total number of particles is constant, Smith defines a Lagrangian multiplier μ such that

$$\frac{\delta}{\delta n} \left\{ E[n] - \mu \int n d\vec{r} \right\} = 0 \tag{101}$$

where $E[n]$ is the total energy expressed as a functional of the particle density, n. The work function can be defined as

$$\varphi_e = - \left[\frac{\partial E}{\partial N} \right] \tag{102}$$

where N is the total number of particles. It can be seen that $\varphi_e = -\mu$. By means of the variational principle and a number of approximations, Smith found a closed form solution for the work function and the potential in the surface region of a one-dimensional metal in terms of the metal density.

Considering that the only parameter Smith introduces into his equations is the metal density, the calculations of surface energy and work functions are surprisingly accurate. Smith's basic model, that of electrons in a positive "jellium" is admittedly crude, and his correlation energy has at best a 20% accuracy claimed for it,[49] but his qualitative predictions must be considered. He finds that if the background positive core charge is considered uniform up to the surface of the metal, there is an overlap of the electron density outside the surface of the metal so that the potential goes continuously from zero at an infinite distance from the metal to a value corresponding to the negative of the work function inside the metal. At large distances from the metal the potential should have the form of the image potential, although this has been somewhat distorted in Smith's calculations by the assumed variational form of the work function.
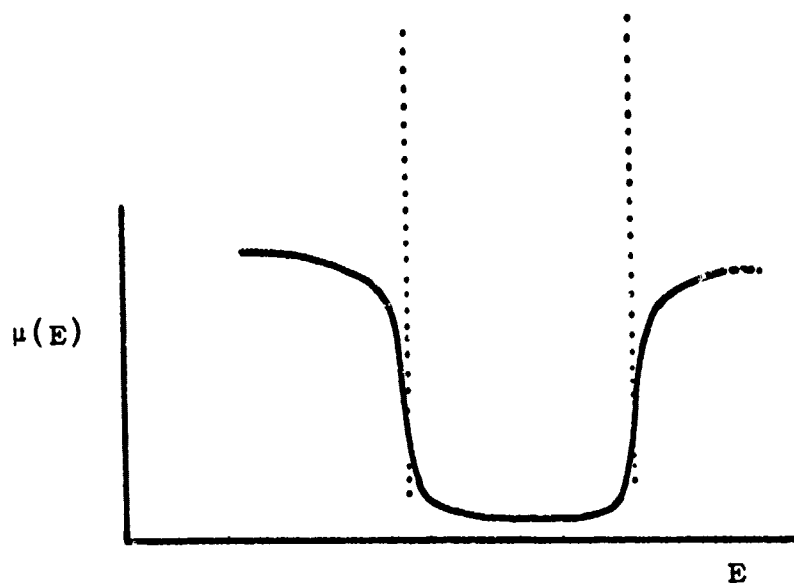
It may be noted that if the contribution outside the metal can be calculated by classical means with a careful treatment of the region in the neighborhood of the metal, then the work function of small particles, such as an isolated metallic sphere, must be increased by a factor inversely proportional to the island radius as shown in Appendix II.

Considering the energy band structure of the substrate, we must first investigate the applicability of using such a model when referring to materials with no long-range order. Theoretical band structure calculations assume one extreme of the Heisenberg uncertainty principle; namely, $\Delta p \to 0$ and $\Delta x \to \infty$. The results of these calculations have been applied down to the region where the spatial localization, defined as the mean distance between scattering, becomes a few lattice spacings. Fortunately, Ziman[50] has found that the short-range order and bonding angle are the fundamental determinants of the appearance of a bandgap. The validity of this model can be seen, for example, in the case of amorphous Si, prepared from the glow-discharge decomposition of silane,[51] which has a room-temperature resistance as much as 12 orders of magnitude higher than the usual type of Si although X-ray analysis shows no structural difference. In the energy band model developed for amorphous materials, it has been assumed that the center of a band shows little effect of the lack of long-range order, as shown by the relatively slight effect of melting on the conduction properties of metals and the optical properties of bandgap materials at energies considerably above E_g, the energy of the forbidden energy gap. At the band edge, however, the sharp transition from the traveling waves in the band to localized waves in the bandgap is replaced by a slow transition to more and more localized states within the bandgap. This is illustrated by Fig. 6a, where the shaded area represents localized states with the very small mobility characteristic of "hopping" conduction. A mobility plot can, therefore, be drawn as in Fig. 6b. Cohen, Fritzsche, and Ovshinsky[52] describe the transition between the high and low mobility states as "mobility edges" leading to a mobility gap rather than a density-of-states gap. High current conduction is thus dominated by a relatively few carriers in the high-mobility localized states.

Assuming that the substrate can be described by the band model, while keeping in mind the band edge describes the mobility edge and that there are localized states extending into the bandgap, the effects of the metal-substrate contacts can be investigated. When the metal is brought into contact with the substrate, electrons will tend to flow from the material with the lower work function to that with the higher work function until the Fermi level in the two materials is equal. This is shown in Fig. 7. The classical picture shown here has several drawbacks, however. If the charge resides in the conduction band of the insulator as a simple wide-band semiconductor and following the suggestion of Mott[53] that most insulators become negatively charged so as to balance the diffusion of electrons from the metal, it can be shown that the extent of the space charge region is given by
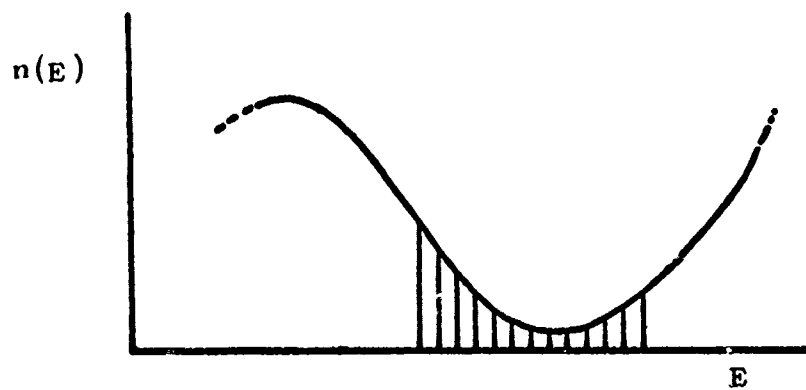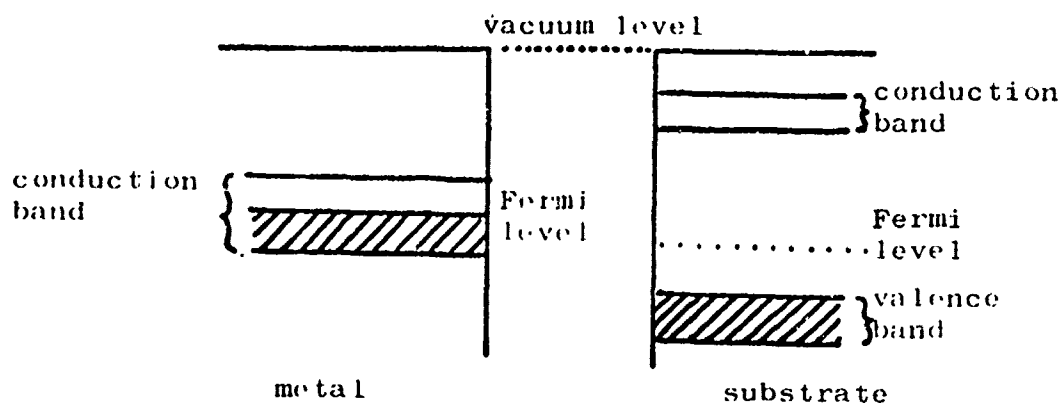
(a) Density of States vs. Energy

(b) Mobility vs. Energy

Fig. 6 - Mobility Gap in Amorphous Semiconductors

44

vacuum level

conduction
band

conduction
band

Fermi
level

Fermi
level

valence
band

metal

substrate

(a) Before Contact

vacuum level

conduction
band

conduction
band

Fermi
level

valence
band

metal

λ
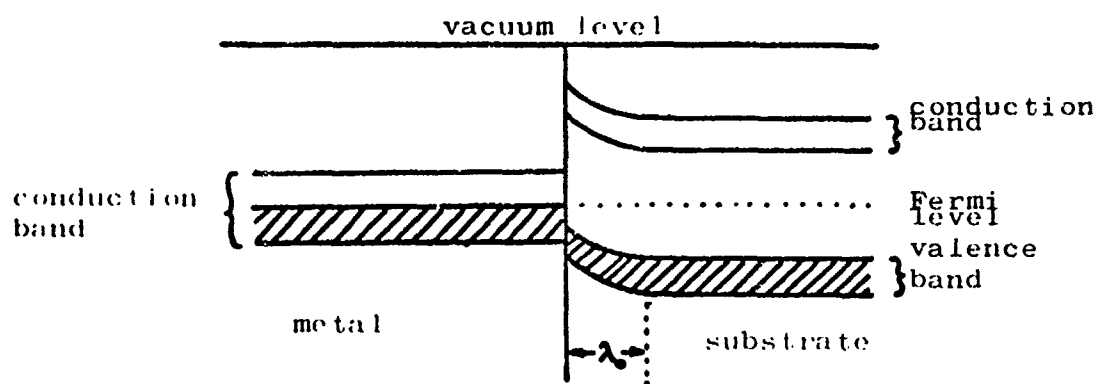
substrate

(b) After Contact

Fig. 1 - Metal-Substrate Interaction on Contact

$$\lambda_o \approx \frac{\pi}{2} \left(\frac{2kT \; K \; \epsilon_o}{e^2 \; N_c} \right)^{1/2} \exp\left(\frac{\psi_i - \chi}{2kT} \right) \tag{103}$$

where N_c is the density of states in the conductor band, K is the relative dielectric constant, ψ_i is the insulator Fermi level with respect to the vacuum level and χ is the insulator electron affinity. This can also be rewritten in terms of the resistivity of the insulator since

$$\frac{1}{\rho} = e_\mu \; N_c \; \exp\left(- \frac{\psi_c - \chi}{kT} \right) \quad . \tag{104}$$

Therefore,

$$\lambda_o \approx \frac{\pi}{2} \left(\frac{1}{e} \; 2kT \; K \epsilon_o \rho \mu \right)^{1/2} \quad . \tag{105}$$

With ρ in ohm-cm and μ in cm^2 per volt-second, this is

$$\lambda \approx 10^{-8} \left(K \rho \mu \right)^{1/2} cm \quad . \tag{106}$$

Fused silica has $K \approx 3.8$ and a volume resistivity of greater than 10^{16} ohm-cm at room temperature, meaning that in a glass slide 1/16-inch thick with a metal contact on one side, a mobility as low as 0.001 cm^2/V-sec will give a depletion layer which permeates almost the entire glass thickness. In fact, the investigation of the dissipation of charge on glass has given values of mobility on the order of 10^{-5} cm^2/V-sec at 80°C.[54] While such a low μ gives reasonable values of λ_o, λ_o would still be quite large compared to the land dimensions. It should be noted that in a nondegenerate carrier gas, the mean free path is given by[55]

$$L = \left(m^* \; kT \right)^{1/2} \frac{\mu}{e} \tag{107}$$

where for low electric fields, T is the lattice temperature and m is the effective mass of the electron in the conduction band. Assuming an effective mass of 0.1 times the free electron mass and T = 300°K, this implies that for μ below 0.1 cm^2/V-sec, the mean free path is less than the interatomic spacing. Friedman[56] has refined this calculation so that the mean free path will be greater than the lattice spacing. This is the only sensible result, since we must have

$$\mu > 0.1 \; W/kT \tag{108}$$

where W is the bandwidth, which can be smaller than kT. In any event, the values for mobility of glass do not satisfy the inequality of Eq. (108) so that the model picture must be modified.

One prime cause for confusion is that the bulk conductivity of most glasses is caused by the diffusion of sodium ions under an electric field.[57] Such ionic conduction gives the same type of exponential temperature dependence as observed in a discontinuous film with an activation energy on the order of 1.3 to 2.6 eV and what is characterized by a very low mobility.[58] Hill[57] obtained the activation energy for the soda glass substrates, which have a large sodium concentration, as 1.4 eV, from which value the band-bending at contact was deduced. It is more likely that the figure actually represents the barrier height for sodium diffusion, which has no direct connection with the electronic processes discussed by Hill.

As has been shown, the mobility must be respectable for the band model to be valid. Therefore, if electrons are injected into the conduction band, these electrons should be able to move in response to an applied external field, unlike the electrons in the valence band. For example, the injected electrons in the conduction band of the insulator diamond have been shown to have a mobility of 1800 cm^2/V-sec[59] at room temperature as compared to a mobility only 35 cm^2/V-sec for copper. This would mean that if such insulators were used to insulate electrodes in vacuum equipment the surface charge acquired from the filament should make the insulator a conductor. This effect does not generally happen, and, in particular, does not happen for the substrates considered, because of localized states acting as traps for the electrons.

One type of localized state, that due to the lack of a regular crystal lattice in an amorphous material, has already been considered. There are two other sources for such states which should be considered: (1) surface states and (2) states caused by adsorbed or dissolved impurities; both of which will be briefly described.

Much theoretical work was done on the consequences of finite boundaries for the crystals during the early development of the theory of solids. Tamm[60] considered a simple one-dimensional semi-infinite series of potential wells bounded by a high potential barrier. This gave an additional energy level within the bandgap whose position was determined by the lattice parameters. Shockley[61] considered the behavior of the energy levels as a row of wells was brought together from a large distance apart, a model somewhat like that considered previously in the case of a metal of finite extent. He found that the surface states appear only when there is band crossing, and then only when the lattice spacing is less than when the bands cross. Various other authors have argued that the free electron model is too general and the extension from one to three dimensions is not straightforward so that too much emphasis should not be placed on these results. Goodwin[62] considered a tight-binding approximation and found that localized surface levels appeared equal in number to the number of surface atoms.

A real surface which is an abrupt termination of the bulk lattice
will have dangling bonds if it is covalent and uncompensated ionic
charge if it is an ionic crystal. In the former case, the LEED investi-
gations of Farnsworth[83] and many subsequent investigators have found a
restructuring of surface, often extending many lattice spacings into the
substrate, which tends to reduce the number of dangling bonds. This is
done, however, at the expense of the bonding angle which must also lead
to disruption of a normal energy band. The exposed ions in the case of
an ionic crystal would most probably attract compensating charge by
adsorbed impurities on the surface.

Impurities can be incorporated into a host lattice in two basically
different ways; they can diffuse into the lattice or they can react with
the surface. Although the following discussion is addressed to the sub-
strate, similar arguments apply to the metal islands. An atom approach-
ing the surface experiences a potential as shown in Fig. 8. The first
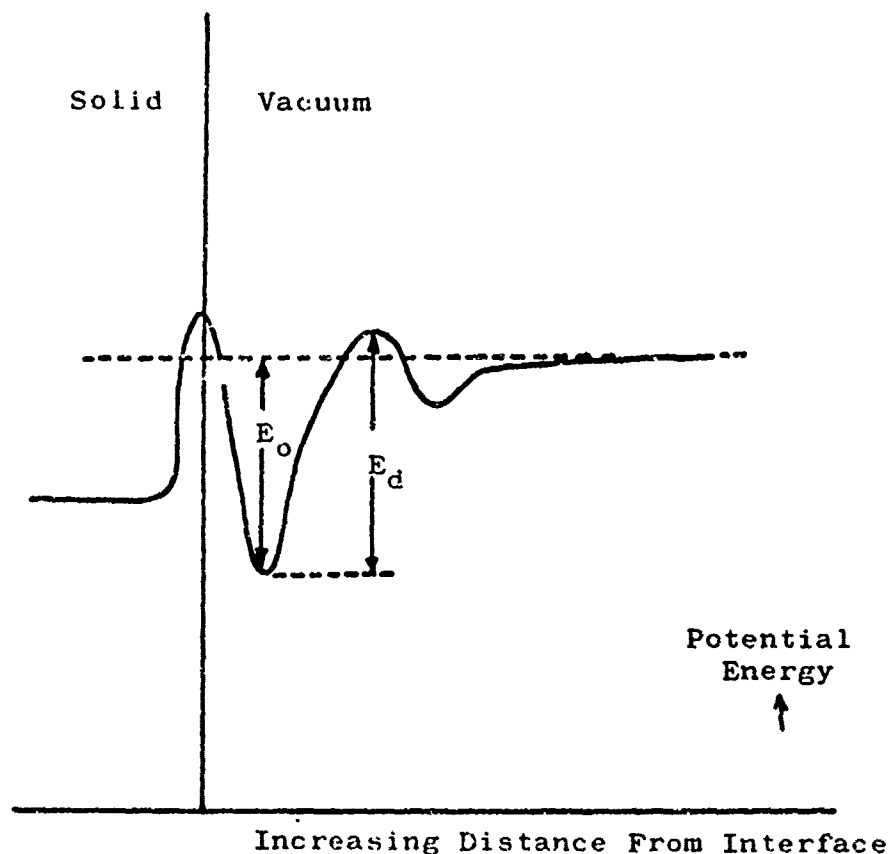potential well is due to polarization and van der Waals forces. An atom

Fig. 8 - Energy Levels of an Atom Approaching a Solid

48

in this well is said to be physisorbed and is bound with energies of less than 0.05 eV. If the adatom is capable of further reaction, the atom can enter the deeper potential well, where it is said to be chemisorbed, which involves an electric interaction. The energy loss suffered by the adatom, the heat of adsorption, E_a, is greater than 0.5 eV and is proportional to the heat of formation of the respective chemical compound. Chemisorbed atoms often require a desorption energy, E_d, greater than the adsorption energy. Sherwood[34] investigated the outgassing properties of clean glass surfaces and found that the sorption of water vapor did not stop at one monolayer, but continued until total amounts equivalent to 10 to 50 monolayers were accumulated. Carbon dioxide was found to accumulate to five monolayers. Pyrex shows considerable structure in the desorption rate for water vapor at elevated temperatures,[35] exhibiting at least two values of desorption energy, E_d, and also evidence of a diffusion-controlled mechanism. Temperature of $400°C$ were necessary to essentially remove all the adsorbed H_2O, but at this temperature the diffusion provides an increased source for further vapor, so that any cleaning process which claims to remove all the water vapor is suspect. Another source of glass-surface contamination is oil-diffusion-pump vapors which are firmly chemisorbed by baked glass surfaces.[66] Cleaning of glass surfaces with hydrofluoric acid has also been shown to result in very persistent outgassing of fluorine.[67]

In addition to the adsorbed species, all commercial glasses contain impurities either purposefully or inadvertently added. Even fused silica can only be refined of metallic impurities at the expense of increased water content.[68] As far as can be determined, no investigation has been made of the impurity distribution within the glass other than for mobile impurities which migrate under external fields such as sodium.

All the impurities, including the adsorbed impurities will have electronic levels which may lie within the bands or in the gaps between the bands. They can also exchange charge with the substrate or the metal islands. The potential magnitude of the charge residing in the chemisorbed water vapor, for example, can be seen from consideration of the static charge transferred by contact charging of gold glass spheres with a contact area on the order of 10^{-7} cm^2.[69] This was found to be as large as the equivalent of 10^{+7} electrons during one contact. Water vapor transfer during the contact was theorized as the charge transfer mechanism. Very large charge transfer was even found in quartz-quartz contact[70] provided different quartz axes were brought into contact on the two spheres.

The preceding discussion supports the picture of many localized states in the forbidden gap in the neighborhood of the surface. Henceforth no distinction will be made between those states arising from the amorphous state of the insulator, the presence of the surface, and the presence of adsorbed and dissolved impurities. Such states modify the picture of the metal-insulator junction as shown in Fig. 7. Instead of forming a double layer by charges in the metal and in the conduction

band of the insulator, charge from the metal will more likely reside
in trapping centers at surface states at the interface or just inside
the insulator. This will result in a much more localized charging of
the insulator, and will imply that the Fermi-level in the insulator and
the metal will line up. This was not the case when the depletion region
extended throughout the insulator.

There is some question in the literature as to the height of the
barrier that would be expected between the Fermi-level in the metal
islands and the bottom of the insulator conduction band. Well away
from the electrode, this should be the separation between the Fermi
level and the conduction band edge characteristic of the substrate
itself. Under the simplified picture of the insulator as a wide-gap
semiconductor and neglecting traps, the Fermi-level should be approxi-
mately in the middle of the gap, assuming almost equal density of states
at the edge of the conduction and valence band. Since quartz has a
bandgap of about 8 eV, this would imply a barrier height of about 4 eV.
In actual practice the Fermi-level would most likely be determined by
the extrinsic levels within the gap introduced by defects and impurities,
and would have to be determined experimentally. Hill[27] reported values
for the barrier height between islands of from 0.5 eV to 2.06 eV,
depending on the metal and island size. He reported values of work func-
tion of 4.5-4.8 eV for the soda glass substrates used. The barrier
height would be the difference between the work function and the elec-
tron affinity of the substrate. Hill's figures, if representative of
the barrier height for wide gaps, would imply electron affinities of
several electron volts. A more direct examination of the metal-SiO_2
interface was made by Deal and Snow,[71] who thermally oxidized silicon
to form an oxide barrier 500 Å thick. The metal-SiO_2 barrier heights
for a number of metal electrodes was measured both by the photoresponse
and from the capacitance-voltage curves for M.O.S. capacitors. For
both methods, Deal and Snow found the barrier height for gold-silicon
dioxide junctions to be 4.1 eV. This value compares with the gold work
function of 4.8 eV and a SiO_2 electron affinity of 0.9 eV. The values
of Deal and Snow are thus much more reasonable than the values often
used in articles on discontinuous films.

It is seen that the energy levels in discontinuous films can be
represented as shown in Fig. 9. Figure 9a is in the plane of the sub-
strate at the film surface. The curvature of the substrate conduction
band between islands is due to the image force attraction to the
islands as illustrated in a simple model by Eq. (A1). A positively-
charged and a negatively-charged island are shown. Figure 9b shows the
energy levels through an island perpendicular to the plane of the film.
An image force-contract potential field is present attracting electrons
to the surface. This will overcome the image-force interaction of an
electron with the dielectric-vacuum interface which will push the elec-
tron deeper into the dielectric where the electrostatic energy is lower.
Between the islands in Fig. 9a there is a large concentration of trap-
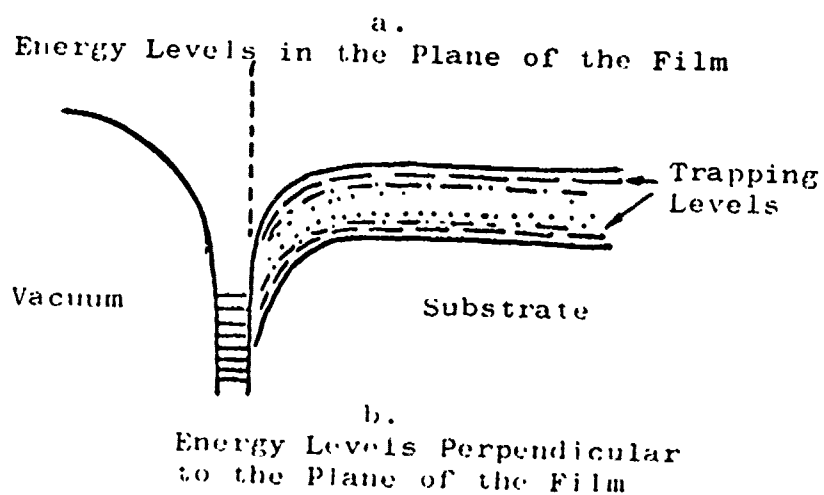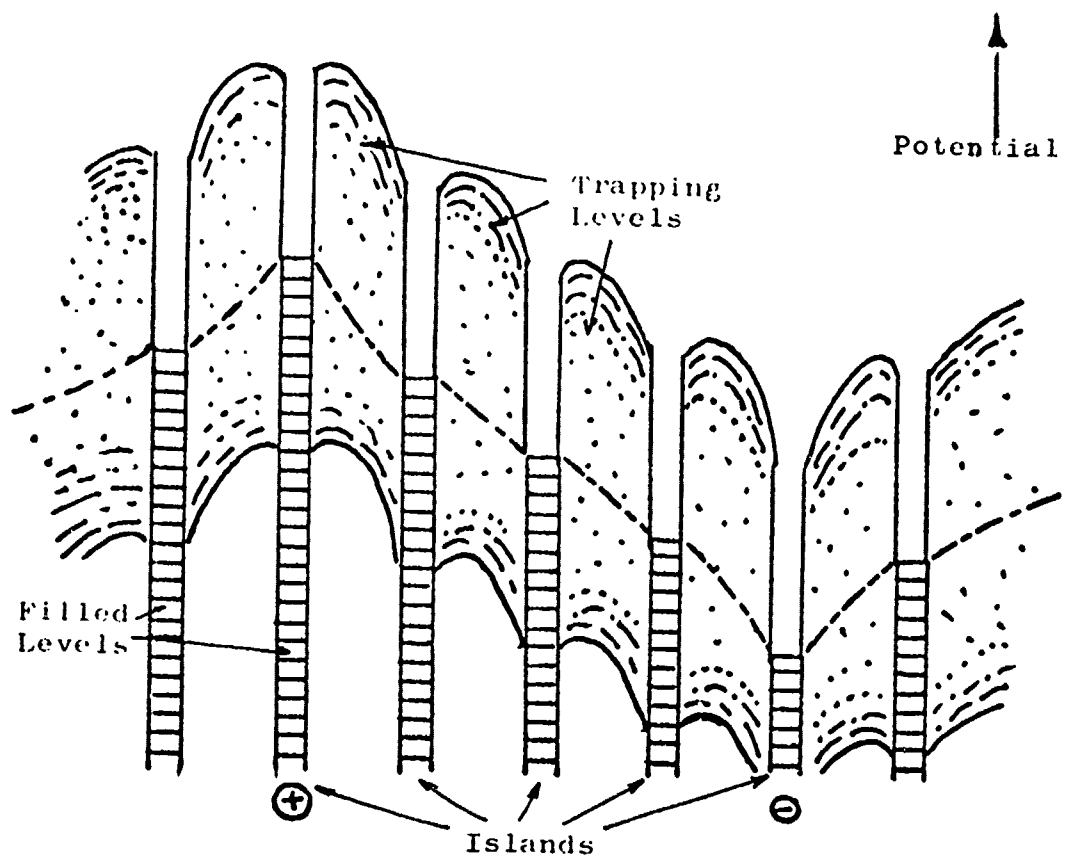ping levels. Figure 9b shows this concentration decreases as one goes

a.
Energy Levels in the Plane of the Film

b.
Energy Levels Perpendicular
to the Plane of the Film

Fig. 9 - Substrate-Island Energy Levels

51

deeper into the substrate. In line with the discussion on metal energy levels, as one goes from the metal to the insulator the potential smoothly joins to that due to the image force, which is the long-range form of the correlation energy.

Islands can be charged positively or negatively by the deficiency or excess of an electron with an activation energy. There is a large density of trapping sites between islands that are quasi-continuous in energy. It is proposed that electron transport between islands is by means of electrons traveling from one island by means of successive tunneling jumps to successive trapping sites to an adjacent island. There are two immediate benefits associated with this model. First, an electron never needs to be raised to the conduction band of the substrate, so that the very high contact potential barrier mentioned previously does not apply to this type of conduction. The second advantage of this model is that allowing the electron to reside at trapping levels between islands allows the dielectric polarization to follow the electron.

It was stated previously that an electron tunneling a relatively large distance must overcome the potential barrier of attraction to the unrelaxed polarized dielectric left behind, as well as accommodate the higher dielectric energy at the new site which has not had a chance to polarize in response to the electron's presence; that is to say, an electron on an island can be considered as a quasi-particle of a type commonly called a polaron. This quasi-particle consists of an electron and an accompanying "cloud" of longitudinal phonons. The general theory of polarons has been much discussed in recent years, and an excellent survey of various approaches to this problem can be found in an article by J. Appel[72] which has an extensive bibliography. One of the features of such quasi-particles is an increased effective mass. This increased mass, by Eq. (62) exponentially decreases the tunneling probability, and in many cases the polaron will have to be excited to its component parts for the tunneling to take place. For small displacements of a polaron, however, mobility is actually aided by the thermal fluctuations of the phonons. There are a number of phonons in the "cloud" surrounding a trap with an electron, creating the polaron. Because the wavelength of the phonons is greater than a lattice constant, the phonons can be considered to be localized in a potential well centered on the trapping site and extending several lattice spaces in all directions. The location of the potential minimum of the phonons will resonate around the lattice site, and when this minimum intersects an adjacent lattice site, the electron can tunnel to this adjacent site with no loss of energy. Thus, a polaron can move across the gap between the islands at a rate slow enough to "drag" the phonon "cloud" along.

Therefore, once an electron from a previously negatively charged island is in the gap between islands, the only barrier that must be overcome is that due to the image force attraction to the surrounding islands. As was shown previously, this image force is reduced from that between planar electrodes. The activation energy for the movement of an excess electron from one island to another island will, therefore,

be overshadowed by the electrostatic energy necessary to put the electron on one of those islands in the first place. The conduction activation energy will most strongly reflect the latter process. Charged island creation can be considered to arise from two sources, which must be examined separately: (1) the process whereby an electron leaves an initially uncharged island and finally is located a large distance away, so that the island that was the source for the electron is positively charged and the island on which the electron finally rests is negatively charged; and (2) the injection of an electron into an island from one of the electrodes.

In the former case when we have charge removed from a previously neutral island, we have the potential around the island the charge has left, as shown in Fig. 2. As explained previously, there are screening effects from the uncharged islands which are not included. In this approximation the potential outside the positively charged island in the presence of an electric field is approximately that given by Eq. (3), resulting in a barrier height of

$$E' = E'_0 - 2\left(\frac{e^.}{4\pi\epsilon_\ell}\, \mathcal{E}\cos\theta\right)^{1/2} \tag{109}$$

where E_0 is the barrier height without the applied field, ϵ_ℓ is the low frequency dielectric constant, and θ is the angle of barrier penetration with respect to the direction of the electric field. The short separation of the trapping sites allows the use of the low frequency dielectric constant. This equation can be written

$$\cdot\,' = 2E_. - \frac{7.6 \times 10^{-4}}{K^{1/2}}\, (\mathcal{E}\cos\theta)^{1/2}\; eV\left(\frac{cm}{V}\right)^{1/2} \tag{110}$$

where K is the low frequency relative dielectric constant. The barrier height $E_.$ would be the electrostatic energy of a single charged island. The factor of 2 arises because the removal of an electron from an originally neutral island creates a negatively charged island on the electron's original site, which is now deficient one electron.

The rate of carrier generation R due to this source is then

$$R = \int_0^{\cdots} \exp\left\{-\left[2E_0 - \frac{7.6 \times 10^{-4}}{K^{1/2}}\, (\mathcal{E}\cos\theta)^{1/2}\right]/kT\right\}d\theta\;. \tag{111}$$

Since this equation has no closed form solution, it is customary to consider the exponent to be dominated by the $\theta = 0$ term so that at high fields

53

$$R \propto \exp\left\{-\left[2E_O - \frac{7.6 \times 10^{-4} \, \text{eV}\left(\frac{\text{cm}}{\text{V}}\right)^{1/2}}{K^{1/2}} \, \mathcal{E}^{1/2}\right]/kT\right\} . \qquad (112)$$

This is the same form as the Poole-Frenkel impurity excitation with which this model has much in common.

The alternative to this type of carrier is to assume that the carrier concentration is determined by the emission rate at the electrodes. In the electrode, an individual electron has negligible electrostatic energy due to the large electrode dimensions. When confined in an island, however, the electrostatic energy E_O, would be required. An electron must, therefore, be given additional energy thermally to be injected into the film. An alternative way of looking at this problem is to consider the film to be a semiconductor in contact with the metal electrodes. In such cases the contact barrier is the difference between the work functions on the two sides of the contact. As shown in Appendix II, if φ_O is the work function of the contact (of the same material as the film), then the work function of an island can be expressed as $\varphi_O + 3/8 \, E_O$. The reason for $3/8 \, E_O$ instead of E_O is the different image force contributions to the work function in the case of the electrode and the island. This energy will be reduced by the Schottky barrier lowering in an applied field, so that the rate of carrier injection into the film is

$$R \propto \exp\left\{-\left[\frac{3}{8} E_O - \frac{3.8 \times 10^{-4} \, \text{eV}\left(\frac{\text{cm}}{\text{V}}\right)^{1/2}}{K^{1/2}} \, \mathcal{E}^{1/2}\right]/kT\right\} . \qquad (113)$$

Here use has been made of the fact that the coefficient of $\mathcal{E}^{1/2}$ is twice as large for Poole-Frenkel emission as it is for Schottky emission.

Equation (113), the one normally used for the barrier lowering at an electrode, assumes a planar electrode with the electron leaving normal to the surface of the electrode. This is not the normal situation encountered in the vicinity of the electrodes. A more realistic model would be a semi-infinite plane conductor with a straight edge, and an electron outside the conductor in the plane of the conductor. This problem can be easily handled by an extension of the work of E. W. Hobsen [] who found the charge induced on a grounded disk of radius, a, at a point Q by a charge e, in the plane of the disk outside the disk at a point P, is given by

$$\rho(Q) = -\frac{e}{2\pi^2} \frac{1}{\overline{PQ}^2} \sqrt{\frac{\overline{CP}^2 - a^2}{a^2 - \overline{CQ}^2}} \qquad (114)$$

where $\overline{PQ}$ is the distance from the charge to the point Q, $\overline{CP}$ is the distance from the center of the disk to the charge, and $\overline{CQ}$ is the distance from the center of the disk to the point Q. As the distance from the edge of the disk to the charge remains constant at d, while the radius of the disk, a, becomes infinite, then the disk becomes a semi-infinite plane of the type that is to be considered. If the y axis is along the edge of the conductor, the x axis is into the conductor so that the external charge is on the x axis at an x value of -d, and the electrode is in the plane z = 0, then the charge density induced in the electrode at a point (x,y) by the charge is

$$\rho(x,y) = - \frac{e}{2\pi^2} \frac{1}{(d+x)^2 + y^2} \sqrt{\frac{d}{x}} \qquad (115)$$

The force on the charge is then given by

$$\bar{F}(d) = \frac{1}{4\pi\epsilon} \int\limits_{x=0}^{\infty} \int\limits_{y=-\infty}^{\infty} \frac{e\rho(x,y)}{(d+x)^2 + y^2} \, dxdy$$

$$= - \frac{e^2}{8\pi^3\epsilon} \int\limits_{x=0}^{\infty} \int\limits_{y=-\infty}^{\infty} [(d+x)^2 + y^2]^{-2} \sqrt{\frac{d}{x}} \, dxdy$$

$$= - \frac{e^2}{8\pi^3\epsilon} \sqrt{d} \int\limits_{x=0}^{\infty} \left[\frac{1}{\sqrt{x}} \cdot \int\limits_{y=-\infty}^{\infty} [(d+x)^2 + y^2]^{-2} \, dy \right] dx$$

$$= - \frac{e^2}{8\pi^3\epsilon} \sqrt{d} \int\limits_{x=0}^{\infty} \frac{1}{\sqrt{x}} \left[\frac{\pi}{2} (d+x)^{-3} \right] dx$$

$$= - \frac{3e^2}{128\pi\epsilon\, d^2} \qquad (116)$$

The potential energy due to this force is

$$E(d) = - \int\limits_{\infty}^{d} F(d')dd' = - \frac{3e^2}{128\pi\epsilon\, d} \quad . \qquad (117)$$

In a constant electric field of value ξ, the barrier for an electron exiting the electrode would be lowered by a factor

$$\wedge E = \sqrt{\frac{3e^{\overline{2}^-}}{32\pi Kc_o}} \; \mathcal{E}^{1/2} = 0.612 \, \mathcal{E}^{1/2} \left(\frac{3.8\times10^{-4} \; \frac{eV-cm^{1/2}}{V^{1/2}}}{K^{1/2}} \right) . \qquad (118)$$

A further complication in these calculations is because the electric field would not necessarily be the applied voltage divided by the electrode spacing. If it is assumed that the deposited film is of negligible conductance, then the field will be enhanced due to the electrode geometry.

The enhancement can be calculated by considering first the case of plane-parallel electrodes, e.g., parallel to the x-z axis at $y = 0$ and $y = \pi$. Here the equi-potential surfaces are planes parallel to the electrodes, and are unaffected if all the space between the electrodes with $z > 0$ is filled with a different dielectric. A conformal transformation from the x,y plane to the w plane can be defined by

$$W = u + jv = a \cosh (x + jy) \qquad (119)$$

The surface $y = $ constant can be considered the equi-potential surfaces of two parallel electrodes $v = 0$, $u < -a$, with potentials 0 and π. Along the surface of the substrate, $v < 0$, the potential between plates V_a at $u < -a$ and 0 at $u > a$ can be written as

$$V = \frac{V_a}{\pi} \left[\cos^{-1} \frac{u}{a} \right] \qquad (120)$$

Thus, the electric field is given by

$$\mathcal{E} = -\frac{dV}{dx} = \frac{V_a}{\pi} \frac{1}{\sqrt{a^2 - x^2}} \qquad (121)$$

which has a maximum in the neighborhood of the electrodes. For idealized electrodes represented by semi-infinite sheets of infinitesimal thickness it can be seen that the electric field immediately adjacent to the electrodes ($u = \pm a$) is infinite. Since the equipotential surfaces were unaffected by dielectric interfaces perpendicular to the electrodes before the transformation, after the transformation the equipotential surfaces, and hence, the electric field, will be unaffected if the region between the electrodes is divided into two dielectric regions by an interface along the plane joining the two electrodes representing the substrate-vacuum interface. If the electrode edge has a radius of curvature h, with h much less than the electrode separation $2a$, then the maximum electric field is approximately that evaluated at

56

$x = a - h$, or a maximum field of $\frac{V}{2a}\left(\frac{2}{\pi}\sqrt{\frac{a}{2h}}\right)$. In any case, the minimum field is smaller than that which would be calculated assuming a uniform electric field between the electrodes by a factor of $2/\pi$.

If this field in the vicinity of the electrodes has the form of Eq. (120) and influences an image force of the form of Eq. (117), then the approximate form of the barrier lowering can be calculated. The potential energy is:

$$\varphi(x) = \varphi_0 - \frac{3e^2}{128\pi K\epsilon_0 x} - \frac{V_a}{\pi}\left[\cos^{-1}\frac{(x-D/2)}{D/2} + \pi\right] \qquad (122)$$

where x is the distance from the electrode, D is the electrode separation, and V_a is the potential applied between the electrodes. Taking the derivative with respect to x and setting this equal to zero, the value of x at the potential maximum, denoted by X_m, is given, assuming $x \ll D$, by

$$\frac{d\varphi(x)}{dx} = 0 = \frac{3e^2}{128\pi\epsilon_0 K X_m^2} - \frac{eV_a}{\pi}\frac{1}{\sqrt{x(D-x)}}$$

$$\approx \frac{3e^2}{128\pi\epsilon_0 K X_m^2} - \frac{eV_a}{\pi}\frac{1}{\sqrt{xD}}$$

$$\therefore X_m = \left(\frac{3eD^2}{128\pi^2\epsilon_0 K V_a}\right)^{2/3} \qquad (123)$$

The lowering of the potential barrier is thus

$$\Delta\varphi \approx \frac{e^{4/3}V_a^{2/3}}{D^{1/3}}\left[\left(\frac{3e}{128\epsilon_0 K}\right)^{1/3}\left[1 + \frac{2}{\pi^{5/3}}\right]\right] \qquad . \qquad (124)$$

This electric field distribution assumes that the current or accumulated charge between electrodes is not large enough to affect the electric field. If the resistance allows a transfer of charge sufficient to build up a significant internal electric field, then between the electrodes there is a current given by the electric field divided by the sheet resistance. Since under steady-state conditions the current is constant at every point between the electrodes, the electric field is inversely proportional to the sheet resistance at that point.

If the resistance of the film should be large over a small segment of the film, the electric field will be high over this region, with a

corresponding decrease elsewhere. Normally, the film's electrical properties would be determined by the high resistance region, unless the field reaches such a value what field emission or tunneling takes place.

For short range trap to trap tunneling, the relative dielectric constant K should be that for low frequencies. Since conduction is close to the surface, however, this may have to be modified. The problem of what value of dielectric constant is applicable to discontinuous film charge transport was investigated by K. van Steensel[?] in a very interesting set of experiments where such films were grown on a barium titanate substrate. Barium titanate is rather unique in that it is a ferroelectric, showing a large spontaneous polarization which undergoes large, sudden changes as the temperature is cycled between room temperature and about -120°C due to the crystal structure reversibly changing from tetragonal to monoclinic to rhombohedral with decreasing temperature. The high dielectric constants are thus very strongly structure dependent and, since they are due to a polarization catastrophe[?] whereby the local field caused by the polarization increases faster than the elastic restoring force of the ionic atoms, causing a polarizing shift in ionic position, the spontaneous polarization is very sensitive to the surrounding medium. In his experiments, van Steensel found no conductivity change which could be attributed to the bulk dielectric constant discontinuity and took this fact to be conclusive proof that the low frequency dielectric constant cannot appear in expressions for the film conductance.

This conclusion is questioned here for two reasons: (1) it has been shown that fast carrier transport using the high frequency dielectric constant required substantial additional activation (see Eq. (70)); and (2) the bulk dielectric constant is probably not typical of the surface. The last statement is because in covalent crystals there is considerable restructuring of the crystal in the vicinity of the surface, and the local field of an atom on the surface is reduced by the lack of the self-aligning forces of neighboring dipoles forming the Lorentz field. In fact, Fig. 10 shows that there must be a polarization field perpendicular to the surface since the field around any closed loop must be zero. This would require that if $\mathcal{E}_2$ is well within the dielectric, $\mathcal{E}_3 < \mathcal{E}_1$, due to the depolarization charge on the ends of the dielectric. Therefore

$$\int(-\mathcal{E}_1 + \mathcal{E}_2 + \mathcal{E}_3 + \mathcal{E}_4)d\ell = 0 \qquad (125)$$

If $\mathcal{E}_2$ is well within the dielectric, then $\mathcal{E}_3$ and $\mathcal{E}_4$ must be non-zero. If $\mathcal{E}_1$ and $\mathcal{E}_2$ are incrementally separated, then $\mathcal{E}_1$ and $\mathcal{E}_2$ must be equal. In this case the local field inside a small cavity just inside the surface of a cubically symmetric crystal is

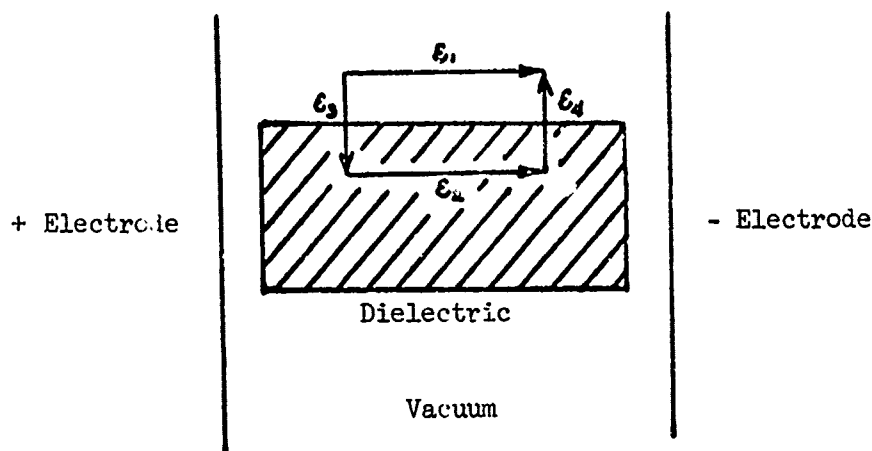$$\mathcal{E}_{loc} = \mathcal{E}_1 \qquad . \qquad (126)$$

Fig. 10 - Electric Field Inside a Dielectric

Deep inside the dielectric, however, with a polarization charge, P, on the ends of the dielectric we would find

$$\mathcal{E}_{loc} = \mathcal{E}_1 + \frac{P}{3\epsilon_0} \tag{127}$$

which would be much larger in the case of materials with a high dielectric constant. Since the polarization in turn is proportional to the local electric field, it can be shown that there is less polarization, and, hence, a lower dielectric constant on the surface of the dielectric. If the charge transfer is localized in this region, as has been proposed for conduction through discontinuous films, this conduction should be less sensitive to changes in the ferroelectric behavior of the bulk material.

EXPERIMENTAL PROCEDURE

Since long term measurements were necessary to follow film developments, an ultra-high vacuum system capable of pressure to 10^{-11} torr has been used. The necessity for this vacuum pressure can be seen by considering the number of particles striking a surface at various pressure levels. The mass of a gas striking a surface is given by[73]

$$M = 58.32 \times 10^{-3} \; P\sqrt{\frac{m}{T}} \; gm/cm^2/sec \qquad (128)$$

where M is the molecular weight of the gas species whose partial pressure is P (in mmHg) and whose temperature is T(°K). At 25°C and atmospheric pressure this implies that 2.9×10^{23} atoms strike a square centimeter of the surface per second, or a monolayer of residual atmosphere is deposited on the surface every 3×10^{-9} seconds. At 10^{-5} torr, which is normally achieved by diffusion pumps, there are 3.8×10^{15} molecules/sec/cm² striking the surface, or a monolayer striking each 0.23 second. Using an ion pump with a baked vacuum system and achieving a pressure of 10^{-11} torr, 3.8×10^9 molecules strike the surface, or a monolayer in 2.7 days. In addition to these considerations, mention should be made of the composition of the residual atmospheres left by the various types of vacuum pumps. In a diffusion pump a cascade of heated oil entraps gas molecules, carrying them from the bell jar area. Better systems have liquid nitrogen-cooled baffles between the oil system and the vacuum to impede any oil vapor diffusing away from the pump. The residual atmosphere in such a system depends on the type of pump oil used and the type of trapping between the pump and the vacuum chamber. By careful attention to all these considerations, pressures down to the 10^{-11} torr range can be attained.[74] It has been found that this atmosphere is composed primarily of those gases not absorbed by the baffling system: neon, hydrogen, and helium in the case of liquid nitrogen traps and zeolite sieves. More common systems include contamination due to backstreaming of the diffusion pump oil through the sieves, especially more volatile hydrocarbons and noncondensable gasses due to the fractionation of the pump oils. During the roughing period there is also a large potential contribution from backstreaming of the oil from the roughing pump. Mechanical pump oil has a vapor pressure of from 10^{-3} to 10^{-5} torr at 50°C. The high rate of gas flow in the roughing line limits this backstreaming. At the end of the roughing cycle this flow is a minimum and often is not sufficient to prevent a significant contamination of the system. Once in the system silicone oils may be polymerized to form a very difficult contaminant to remove. Water vapor is also found in many diffusion pumped baffled systems since there is a significant vapor pressure of water over ice condensed on the

Preceding page blank

baffle. These adsorbed impurities can influence the leakage current on the substrate surface, the adherence of deposited films and the mobility of atoms and nuclei diffusing across the surface.

To avoid these problems in the experiments to be described, the oil-free ultra-high vacuum system shown in Fig. 11 was used. The main pumping elements were four triode sputter-ion pumps. Since the pressure must be on the order of 10 um for the initiation of the glow discharge in this type of pump, a two-stage cryogenic pump was first used to reduce the pressure from atmospheric. Each stage of the fore-pump consisted of zeolite beads, characterized by an extremely large surface-to-volume ratio due to an open lattice structure, in thermal contact with a liquid nitrogen reservoir. At liquid nitrogen temperatures, most gases either condense on or are cryotrapped by the zeolite. Hydrogen and the noble gases are not easily removed by zeolite, so that the residual atmosphere was rich in these elements, even though they form a small percentage of the atmosphere. After each pumpdown the zeolite was rejuvenated by heating to 150°C to drive off the adsorbed gasses and water vapor. After the pressure was reduced by the cryogenic pumps in the bell jar to about 10^{-2} torr, the bakable main vacuum valve was closed and titanium filaments resistance heated to evaporate titanium onto the bell jar walls. At room temperatures titanium chemisorbs many common gasses. In particular, it can form stable oxides, carbides, hydrides, and nitrides, so that O_2, CO_2, CO, CH_4, H_2O, and N_2 are strongly bound up to a monolayer in thickness on the surface of freshly deposited titanium. When the titanium surface was continually or intermittently refreshed by re-evaporation, very high pumping speeds were possible. For example, as small an evaporation rate as 0.1 gram per hour allowed a pumping rate of 60 torr-liter sec^{-1} for H_2, 4.5 torr-liter sec^{-1} for O_2 and 0.5 torr-liter sec^{-1} for N_2 and H_2O. Noble gases, because of their chemical inertness, were only slightly pumped by this method.

When the pressure within the vacuum chamber had been sufficiently reduced, preferable below about 5 um, the triode sputter-ion pumps were turned on. These pumps were modifications of Penning vacuum gauges, which pump in their own right. A very high electric field was established which draws a slight electric current from the cathodes. A high intensity magnetic field parallel to the electric field forced these electrons into a helical path, increasing the probability of collision with a residual gas molecule. When such a collision took place the gas molecule was ionized, and accelerated toward the cathodes. In triode pumps such as have been used in the experiments to be described, the cathodes were of a honeycomb-structure so that most of the gas molecules passed through without striking the cathode. They were then decelerated and strike collectors of the same metal as the cathode (generally Ti) which were at a higher potential than the cathodes. The gas atoms were generally chemisorbed, having too little energy to cause sputtering. Those gas molecules which struck the cathode, generally at an oblique angle, sputtered material from the cathode unto the collector. This
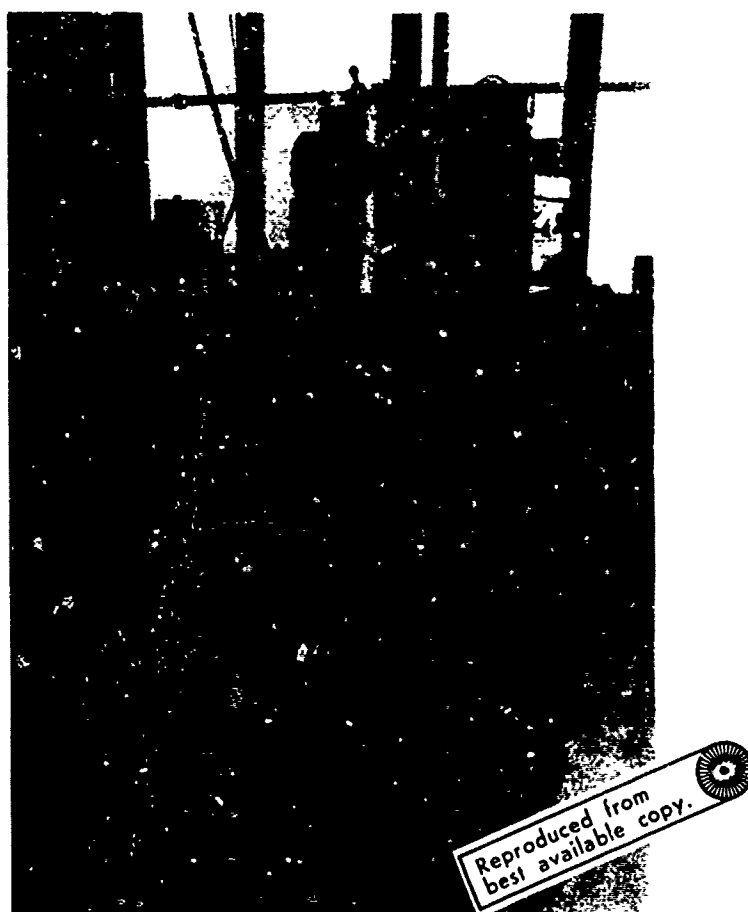
Fig. 11 - Vacuum System

continually renewed the collector so that the surface did not become saturated by the adsorbed gases. Since the gases were thus continually buried beneath new metal layers on the collector, a strong chemical bond was not necessary to hold the gas on the collector. Thus, reasonable pumping speeds could be obtained for noble gases with this type of pump forming a good complement to cryopumping and sublimination pumping which are not effective with noble gases.

In order to take full advantage of the pumping ability of the system used, it was necessary to desorb gases from the walls of the bell jar. The bell-jar walls were heated to 250-300°C with the upper portion of the bell jar enclosed in an insulated oven containing resistive heating elements. These elements were powered by a 40 ampere triac driven by a zero voltage switch. Temperature control was achieved by a diamond thermistor in a resistive bridge controlling the zero voltage switch. As an added protection against control failure the maximum power output was limited by the period of an adjustable neon-bulb oscillator, and the bakeout time automatically controlled by the discharge time of a coulometric cell connected to the gate of the zero voltage switch. The lower part of the vacuum chamber, containing the pumps, was not directly heated since the Curie temperature of the magnets of the ion pumps used was about 250°C. Heating the magnets above this temperature could have caused permanent loss of magnetism. Thermal conduction through the bell jar raised the temperature of the base over 100°C, which was not sufficient for an efficient bakeout. This deficiency was somewhat mitigated by the fact that the titanium sublimation pump was directed so as to deposit the titanium film over the inside surface of the bottom of the vacuum chamber. Adsorbed gases which were not desorbed by the bakeout in this region of the bell jar could be "buried" by a layer of freshly evaporated titanium.

Evaporation of the gold films was accomplished from a resistively heated molybdenum boat. The film was deposited on a quartz substrate with current electrodes formed from chronium thin films overlaid with gold. Figure 12 shows the structure of the substrate. The spacing between the electrodes was normally 0.01 cm and was determined by using a wire mask during the electrode deposition. Contact to the electrodes was made by a thermocompression bond between 0.0126 cm diameter gold wire-balled on the end, a small gold pad and the gold-on-chromium electrode. This combination had been found to result in excellent mechanical strength and low contact resistance. On the same substrate a thermocouple formed from a 0.01 cm chromel wire and 0.01 cm alumel wire welded to a 304 stainless steel pad was attached to the substrate by means of Vacseal.[78] The substrate was spring loaded against the cold trap with a tin pad sandwiched between the two surfaces. During the bakeout the tin melted to make good thermal contact between the substrate and the cold trap. This cold trap had been constructed from 0.0126 cm thick stainless steel tubing of $1\frac{1}{2}$-inch diameter heliarc welded at the top to a Conflat feedthrough and at the bottom to a large thermal capacity stainless steel block. The thin-walled tubing served as a low thermal conductivity reservoir for liquid nitrogen to lower the temperature of the substrate. A heater coil could be lowered to the stainless steel.

64
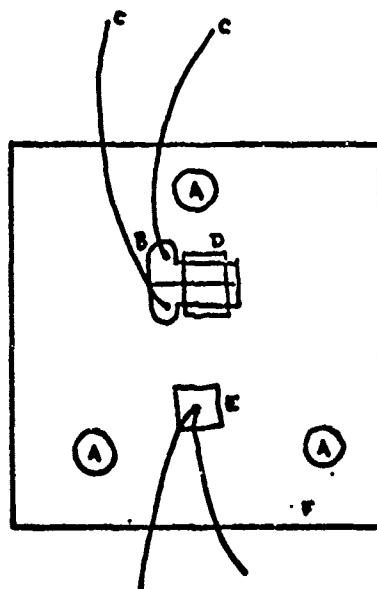
A Masked depositions for electron microscope study

B Gold-on-chromium electrodes

C Gold wires making thermocompression bond connection with the electrodes

D Mask-defined deposition for the study of current effects

E Chromel-alumel thermocouple welded to a stainless steel pad bonded to the substrate

F Glass substrate

Fig. 12 - Substrate Layout

block to raise the substrate temperature. The Conflat flange also had four insulated kovar feedthroughs for the electrical connections. A mask on the substrate holder confined the film deposition to a controlled area on the substrate.

Film deposition was monitored by observing the resistance of the film and by observing the change in the resonant frequency of a quartz crystal when the evaporant condensed on one face of the crystal. This method of monitoring depositions was first investigated by Sauerbrey[75] and Lostis.[76] Considering the major surfaces of the quartz wafer of thickness d_q as antinodal, the frequency f for thickness-shear oscillations is

$$f = \frac{c_t}{2d_q} = \frac{1.67 \times 10^3 \ \text{Hz mm}}{d_q} \tag{129}$$

where c_t is the propagation velocity of the elastic wave in the direction of thickness. An AT cut quartz crystal was used to take advantage of the fact that in a temperature interval about $\pm 30°C$ about room temperature the temperature coefficient of the frequency is smaller than $\pm 5.1 \times 10^{-6}$ deg^{-1}. With a mass M of deposit over an area Am on the quartz, it can be shown[77] that the change in frequency Δf is given by

$$\Delta f = -6.3 \times 10^4 \ \text{Hz} \ g^{-1} cm^4 \ \frac{K \cdot \Delta M}{d_q^2 A_m} \tag{130}$$

where $K \approx 1$ is a factor reflecting the distribution of the deposited mass ΔM. The circuit used to determine the change in frequency is shown in Fig. 13.

Inside the vacuum the crystal was held by a water-cooled, spring loaded holder in a position adjacent to the substrate. It had been found that during the deposition, the quartz crystal became inoperable since radiant heat from the deposition source introduced strains in the crystal which gave rise to spurious modes of oscillation. For this reason the crystal oscillator was used only to determine the total amount of gold laid down by measuring the change in frequency before and after a deposition

The general procedure used during a deposition was to set the crystal oscillator voltage output to zero after a vacuum of at least 10^{-10} torr was obtained as measured by a cold cathode discharge gage. With a movable shutter covering the boat, the evaporation boat was brought up to its final temperature and regulated by controlling the boat current. Two boat designs were used during the course of the experiment; one, a 0.0127 cm thick molybdenum dimple boat one-half inch wide and four inches long, and the other, a 2.5 cm by 1.80 cm boat of 0.005 cm molybdenum welded to tantalum electrodes forming an integral heat shield. The
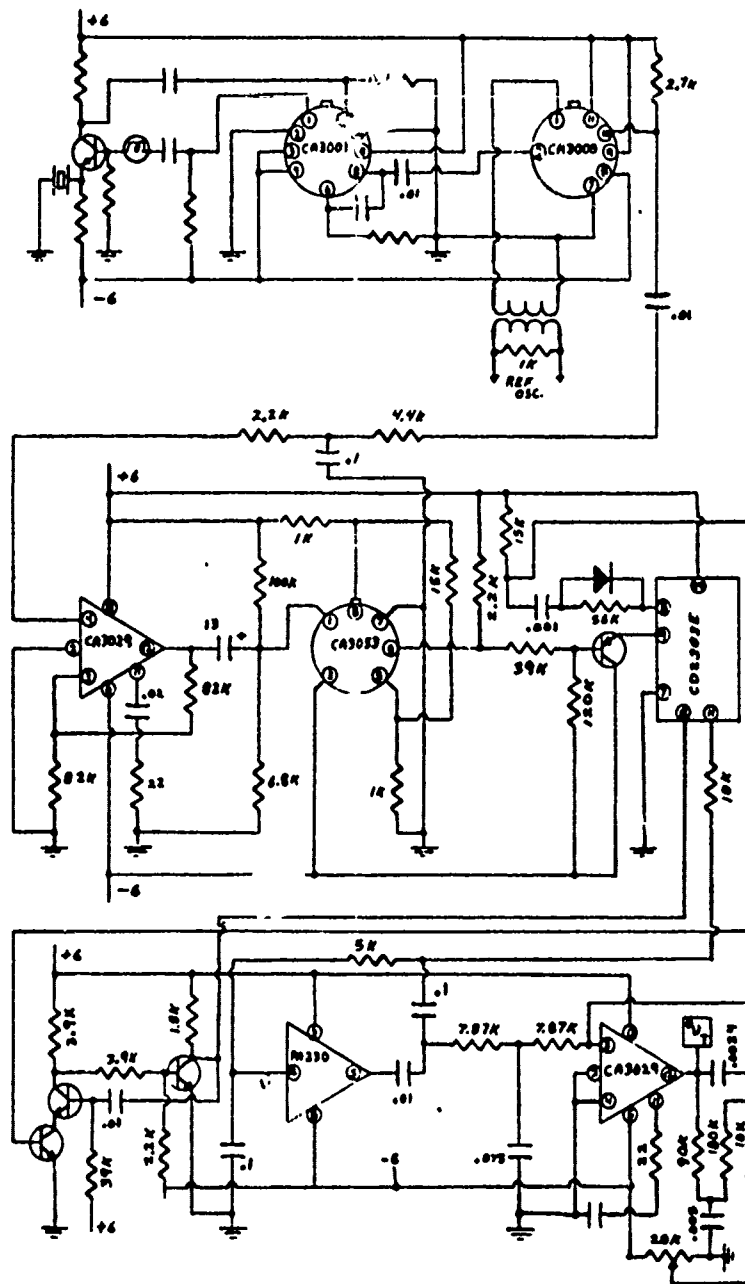
66

Fig. 13 - Crystal Oscillator Circuit

67

current supplied to the first boat was generally 35 A and that supplied to the second boat was 80 A. The shutter was then opened and the desired amount of film, as determined by monitoring the resistance between the electrodes, was deposited on the substrate. Current to the boat was then stopped and the shutter again closed and measurements of the deposited film's resistance initiated. Pressures during deposition were usually on the order of 10^{-9} torr, dropping to below 10^{-10} torr soon after the deposition was stopped, with all resistance measurements made below this last limit. Either the resistance was monitored as a function of time at a constant current and voltage, or the substrate was cycled over the temperature range of interest as rapidly as possible and values of the current through the substrate recorded at representative voltage levels at many temperatures in order to determine the resistance as a function of temperature and voltage. In order to avoid thermal shocks to the system and to limit thermal gradients which would lead to inaccurate temperature measurements, the temperature cycling was limited to approximately a three-hour period.

After sufficient data was taken, the deposition step could be repeated to investigate the effect of the addition of more gold on the conduction processes. When the final data was taken the system was back-filled with dry nitrogen and the substrate removed to another vacuum system where the substrate was coated with a semi-transparent layer of carbon. The substrate was then immersed in a dilute hydrofluoric acid etch (about one drop of HF in 10 cc distilled water). The carbon and the gold film on a test area of the substrate then floated free from the quartz substrate, and could be caught on copper grids. These grids were then mounted in a Philips 300 transmission electron microscope to examine the structure of the original films. Both transmission electron micrographs and electron diffraction photographs were taken.

EXPERIMENTAL RESULTS

Three main features of nucleating films were investigated: (1) the
influence of temperature on the electrical properties of such films,
especially the relation between activation energy and film structure;
(2) the influence of electric field on film electrical properties; and
(3) the influence of annealing on the electrical properties of such
films. In addition, a new breakdown phenomena was encountered and
investigated, and preliminary work was done on the influence of elec-
trode spacing on the electrical properties of the films.

It is well known that discontinuous films show a negative tempera-
ture coefficient of resistance, with resistance decreasing almost expo-
nentially with absolute temperature. This is opposed to the positive,
almost linear, temperature coefficient of resistance found in bulk
metals. Figure 14 shows a plot of the response of one particular film
which is representative of the general behavior. It can be seen that
if the current is considered to be proportional to $\exp(-E_a/kT)$, where
E_a is an activation energy, k is Boltzmann's constant and T is the abso-
lute temperature, as is usually done when considering a thermally acti-
vated process, then various activation energies can be assigned to a
particular film depending on the temperature and the applied voltage.
Figure 15 shows a plot of the activation energy (calculated from the
straight lines shown in Fig. 14) as a function of the square root of
the applied voltage. A photograph of the film from which this data is
taken is shown in Fig. 16. It should be noted that the portion of the
film in the micrograph was not taken from between the electrodes, but
from a separately masked area adjacent to the electrodes. Although the
initial films should therefore be identical, it will be shown that an
applied field affects the film structure, and thus the deposit between
the electrodes may be modified during the resistance measurement. It
is hoped that the small voltages used in an ordinary resistance measure-
ment would cause these changes to be negligible. This is further sup-
ported by the reproducible resistance curves over short time periods,
before annealing effects become evident.

From Figs. 15 and 16, and Figs. 17 and 18, which are from another
film, it can be seen that a given set of data exhibits both a range of
island sizes and activation energies, so that an exact correlation of
the two is difficult. The experimental data can be correlated with
Eqs. (1), (2), (17), (28), (29), or many other forms of these equations
which have been presented in the literature. This is especially true
since the value of K in these equations is somewhat indeterminant. One
complicating factor that has been neglected in the theoretical consider-
ations is the distribution in island sizes and shapes in an actual film.
This reality is the most likely source of the reduction in activation
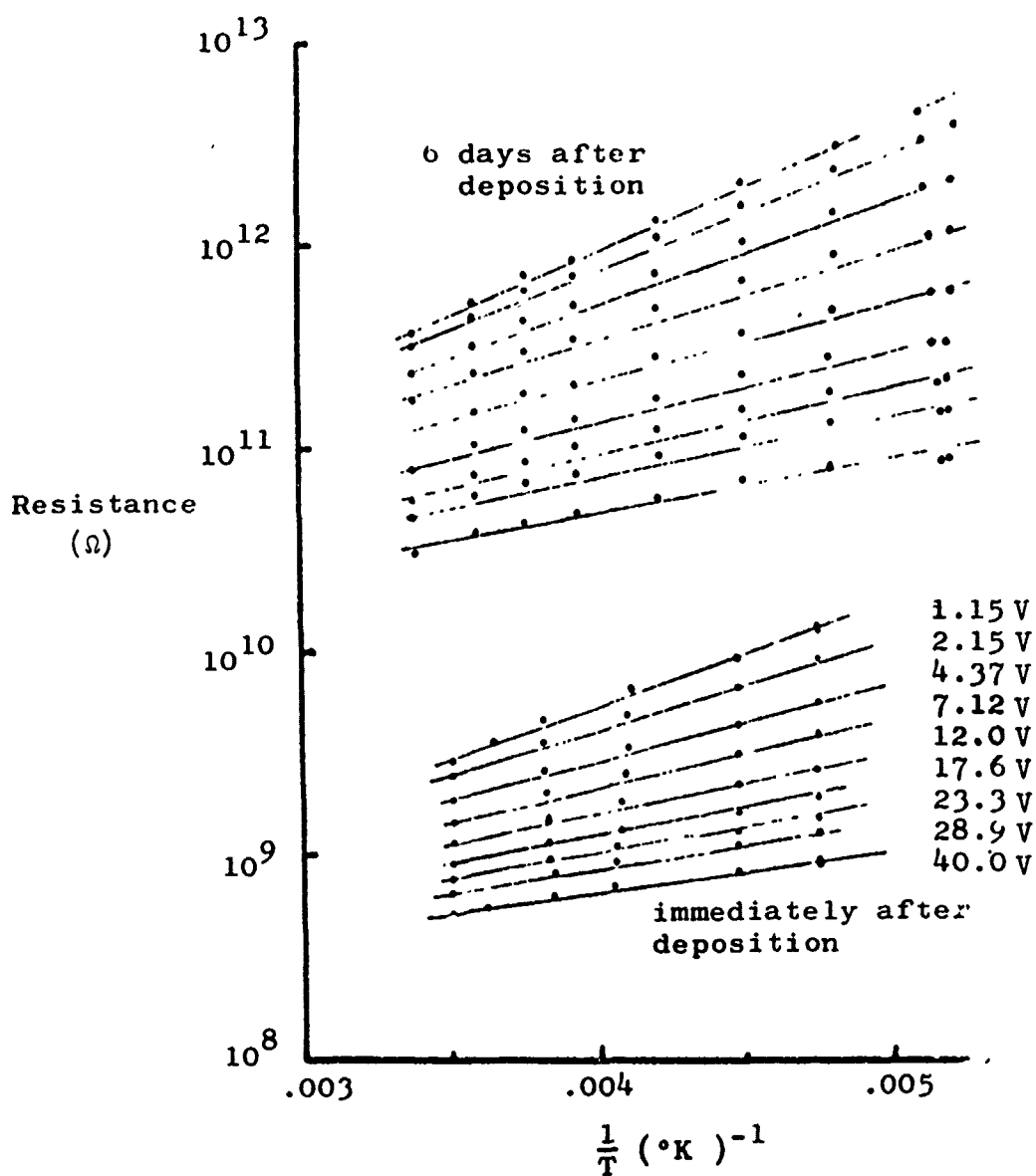
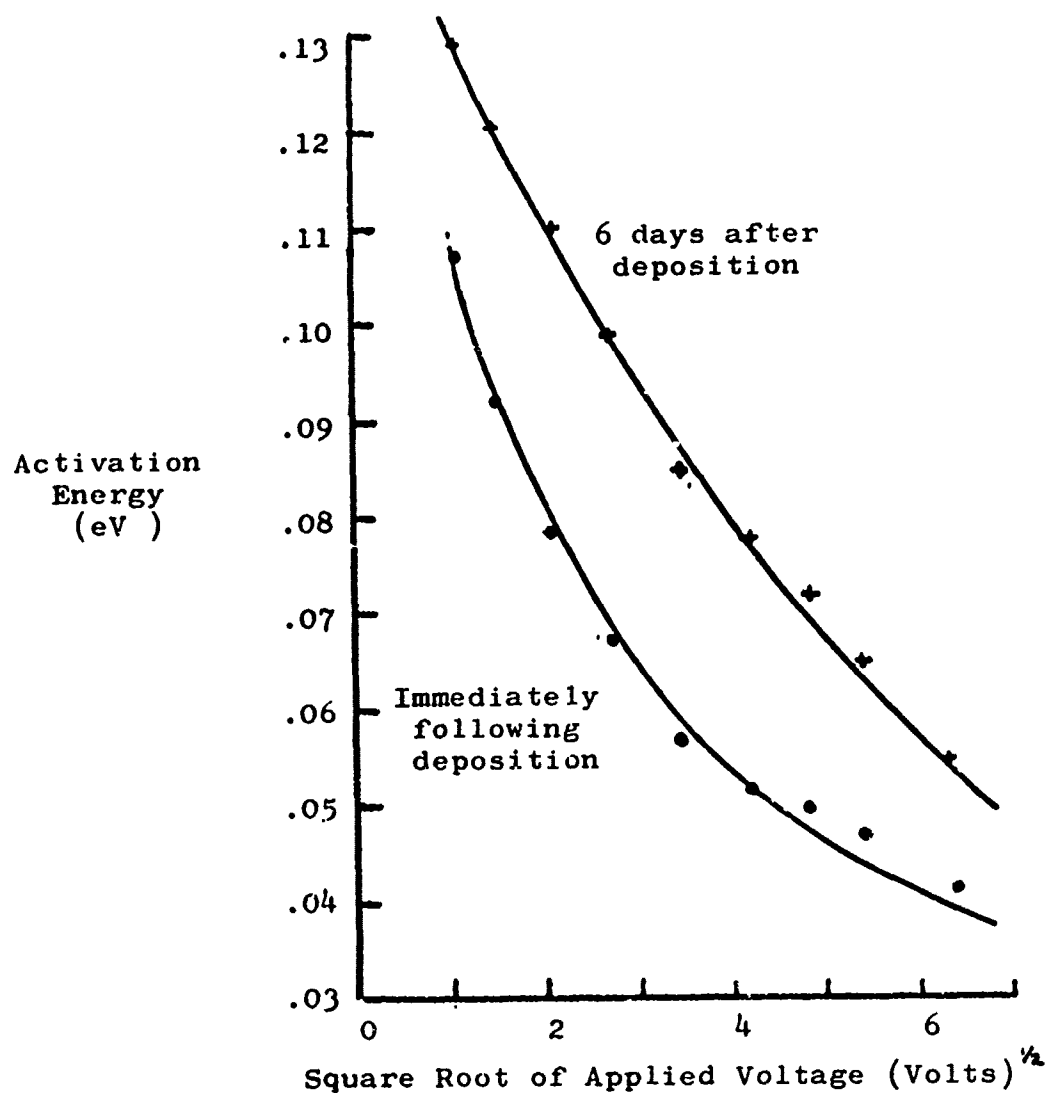Fig. 14 - Film Resistance as a Function of Temperature

1000 Angstroms

Fig. 16 - Electron Micrograph of Film

Fig. 15 - Time and Applied Voltage Effects on Activation Energy

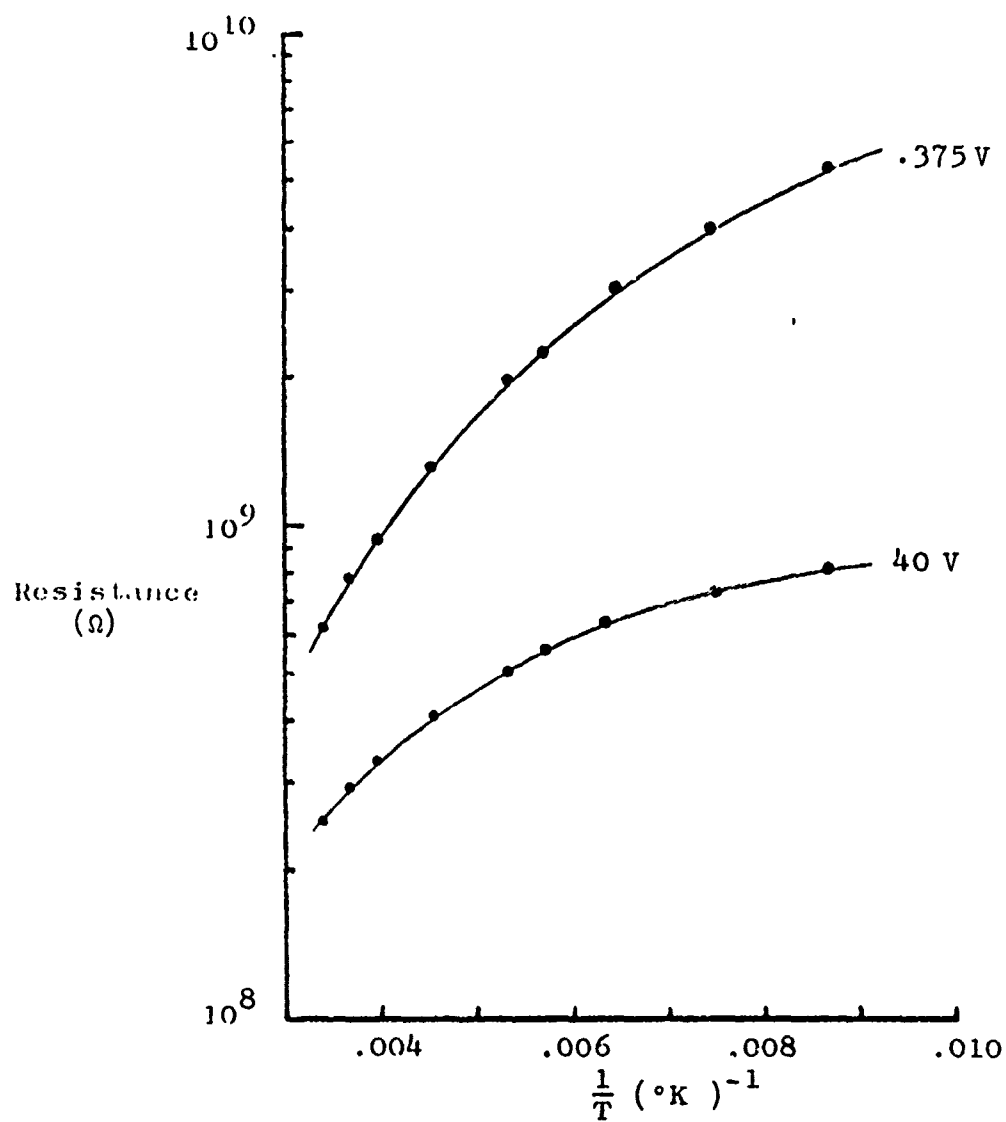Fig. 17 - Film Resistance as a Function of Temperature
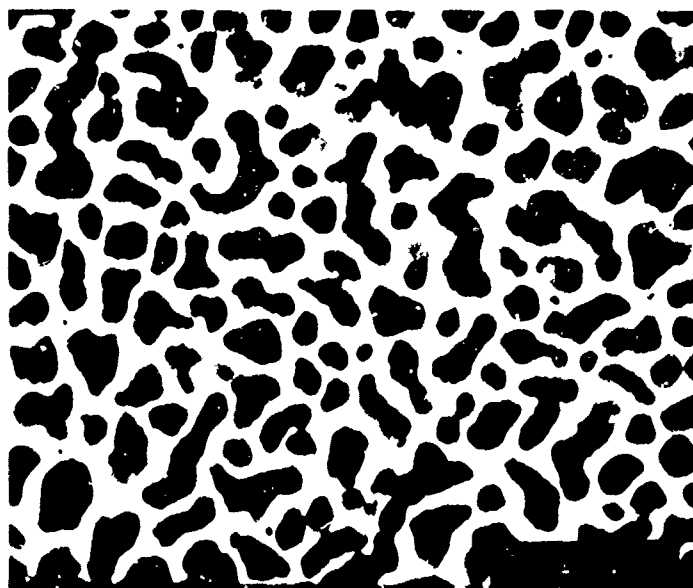
1000 Angstroms

Fig. 18 -- Electron Micrograph of Film

74

energy calculated at decreasing temperatures with a constant voltage.
As an example, we could consider the effect of a film consisting of two
island sizes: one island size with an activation energy of 0.016 eV,
and smaller islands with an activation energy of 0.9 eV, which are forty-
two times as numerous as the larger islands. Since the theoretical dis-
cussion has shown that the most significant dimension of a disk-shaped
island is the major semiaxis, it can be seen from Fig. 18 that the model
using two island sizes, one forty-two times as numerous and about 0.18
the size of the second, is a reasonable crude approximation of the film
represented here. With this mathematical model, an activation energy,
E_a, defined as

$$E_a = \left(\frac{1}{T_2} - \frac{1}{T_1}\right)^{-1} K_b \, \ell n \, (n_1/n_2) \tag{131}$$

where

$$n_n = C\left[e^{-\frac{0.016 \text{ eV}}{K_b T_n}} + 42e^{-\frac{0.9 \text{ eV}}{K_b T_r}}\right] \tag{132}$$

can be constructed. Here C is a constant which is assumed to be inde-
pendent of T, or slowly varying with respect to the exponential functions
of temperature, and k_b is Boltzmann's constant. The activation energy
so defined is plotted in Fig. 19 as a function of 1/T, and three points
from the data shown in Fig. 17 for a voltage of 0.375 volt where field
lowering of the activation energy would be minimal. It should not be
inferred from the relative agreement of the experimental points that the
assumed model is more than a very crude representation of the actual
film since there are an infinite number of families of curves of $N(E_a)$,
where N is the density of islands per square area with an energy in a
range dE_a about an activation energy E_a, which would duplicate the curve
shown. In general, however, it can be said that the curvature of the
plots of the logarithm of the current through the film or the resistance
of the film at a constant voltage is related to the uniformity of the
island size.

Another interesting aspect of the film resistance is the nonlinear
current-voltage relation of discontinuous films. Many previous investi-
gations of such films have reported a film resistance which is constant
at low fields and at higher fields shows a logarithm of the resistance
which is proportional to the square root of the applied voltage. Such
behavior was observed in the films investigated during the study,
although for the film thicknesses used, the high fields resulting from
the small electrode spacing limited the linear region which could be
observed. Figures 20 and 21 show the trend toward a constant resistance
at low fields and the $v^{1/2}$ dependence at large fields for two different
films. As can be seen in Fig. 11, this film showed no deviation from
the exponential dependence of the resistance on the square root of volt-
age up to an applied voltage corresponding to a uniform field of 100 kV/cm.
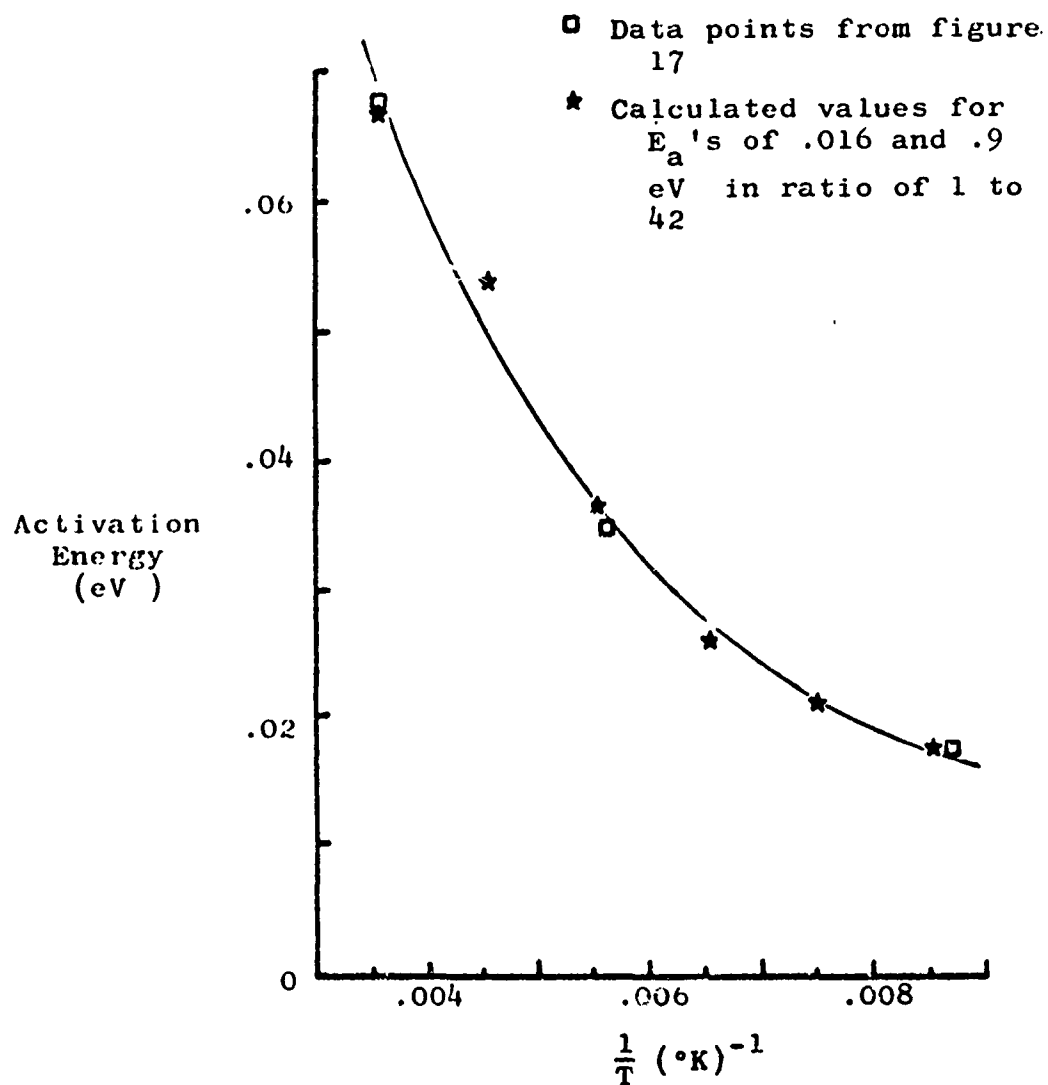It should be noted that the high-field resistance-voltage curve shown

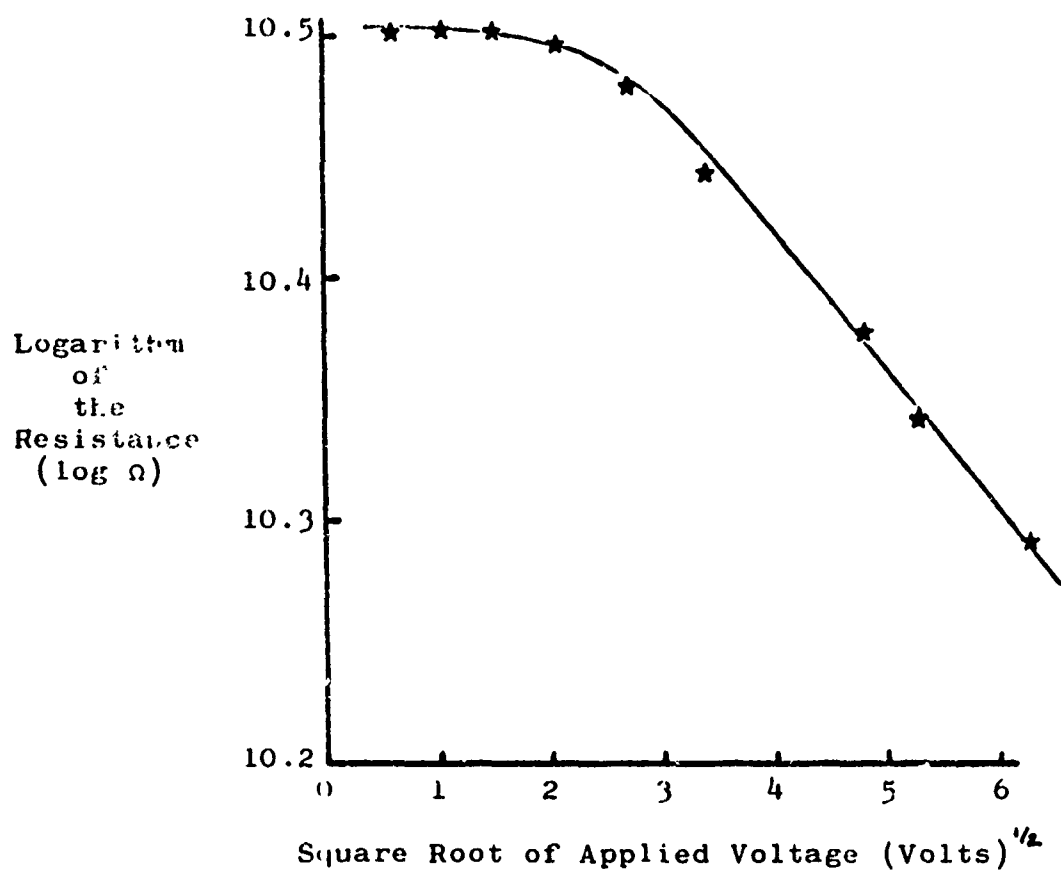Fig. 19 - Activation Energy f· : land Size Distribution

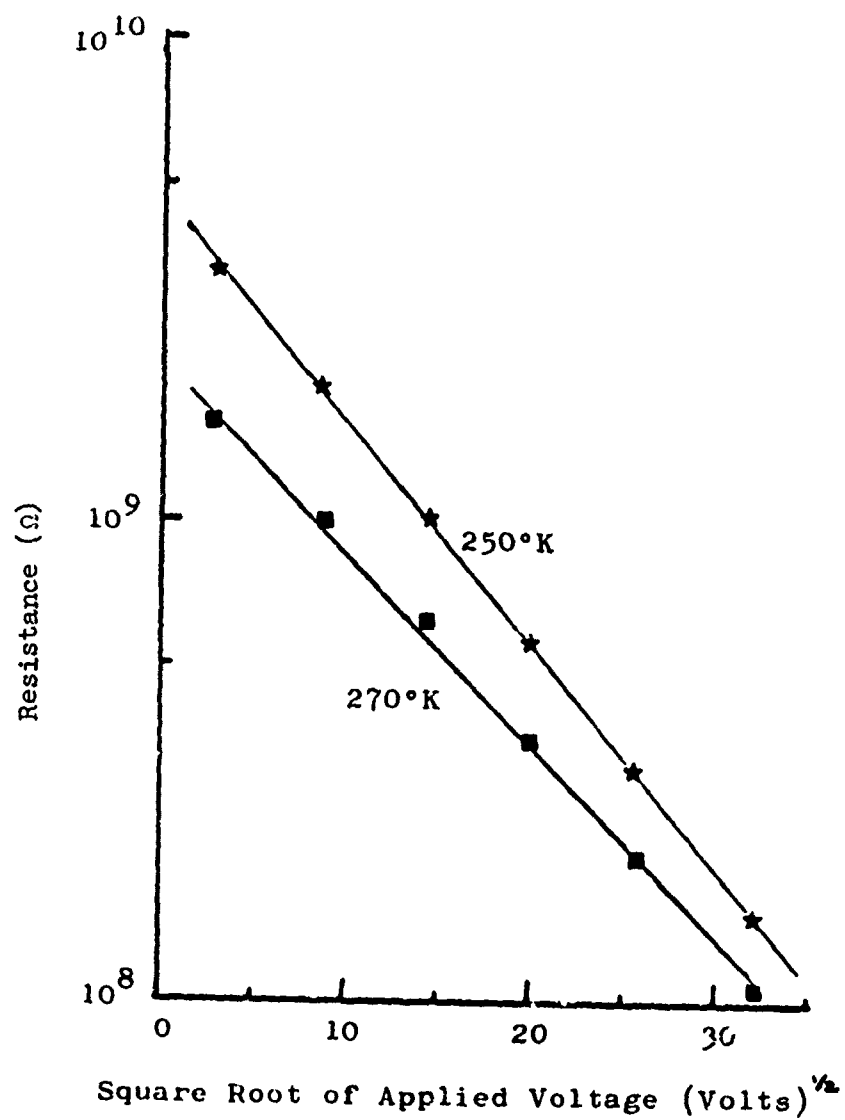Fig. 20 - Resistance as a Function of Applied Voltage

Fig. 21 - Resistance as a Function of Applied
Voltage

was not always encountered, possibly due to resistive heating of portions of the film. When the applied field causes a lowering of the activation energy E_a, and the electric field has a value ξ, then the barrier lowering can be expressed as

$$\frac{dE_a}{dv^{1/2}} = \frac{\beta}{K^{1/2}} b \tag{133}$$

where K is the relative dielectric constant, b is defined by

$$\xi = b^2 \frac{v}{D} \tag{134}$$

where v is the applied voltage, D is the electrode separation. β has a value of 0.0076 eV/$v^{1/2}$ in the case of a film-determined barrier lowering, a value of 0.0038 eV/$v^{1/2}$ for Schottky barrier lowering at the electrodes, and a value of 0.0023 eV/$v^{1/2}$ in the case of a flat electrode seen on edge. For all cases $b \geq 1$ since more energy is required to create charges than to move them once created. Hence, carrier generation, which is assumed to determine E_a, is dominated by the high-field regions of the film.

There are several nonequivalent ways of experimentally investigating the effect of applied fields. For example, while a plot of log R vs. $v^{1/2}$ generally yields a straight line, log I vs. $v^{1/2}$ does not. This requires a consideration of which model one is taking for the conduction mechanism. The current can be represented as the product of a supply function proportional to $\exp(-E_a/kT)$ and a transmission function which shows the ease with which . charge can be moved. The transmission function can be written as the difference between forward and backward transmission over a barrier, expressed as

$$D \propto [\exp(+\eta) - \exp(-\eta)] \propto \sinh \eta \quad . \tag{135}$$

In this expression

$$\eta \propto (e\, \xi_x - \delta E) \tag{136}$$

in the case of tunneling, with ξ_x the applied field and δE the field-induced barrier lowering. In the case of thermionic emission

$$\eta \propto e\, \xi_x \tag{137}$$

Equation (136) is based on Simmons approximation of the WBK equation.[37] For small values of η, such as are normally encountered, Eq. (135) becomes almost linear in ξ, and, hence, in V, if the field is constant

79

across the interelectrode spacing. This would yield the linear log R vs. v" plots if the activation energy is lowered by a term proportional to t ", as has been assumed. The constant of proportionality in Eqs. (1 35), (1 36), and (137), together with the fact that $\mathcal{E}$ is not necessarily proportional to the applied field, complicates the evaluation of the coefficient of $\mathcal{E}$ ' in the exponental. This can be seen from the fact that evaluation of d (log R)/dv resulted in values of from 7.5 to 51.5 as compared with theoretical values of 39.2 b/K for bulk "Poole-Frenkel" type barrier lowering and 12.0 b/K for edge emission from the electrodes, where b is defined by Eq. (134). A plot of log I vs. v does not give a straight line. A straight-line approximation to the curve results in values of between 23.2 and 125.

The more direct calculation would be to evaluate the activation energy from the log I vs. 1/T plot, and determine the change in these values with the square root of the applied voltage. For typical films the calculated values are given in Table I. Several features of this table are noteworthy. If the field were uniform across the interelectrode space the variations in dE_a/dv observed should not be seen. The largest values of dE_a/dv are associated with very high resistance films. While the experimental errors in the current measurement are potentially greater when small currents are considered, the results from the two films which, like that shown in Fig. 8, showed decreasing activation energy with decreasing temperature also show an increased dE_a/dv at lower temperatures and, hence, lower currents.

Table I - Voltage Reduction of the Activation Energy

E_a (0.375 V) (eV)	I (0.375 V, 300°K)(amp)	(eV/v)
0.1 3	8.6 x 10⁻¹	-0.013
(1) 0.0177	5.8 x 10⁻¹⁰	-0.00175
(1 0.066		-0.00378
0.00	8.82 x 10⁻¹¹	-0.00228
0. 99	8.7 x 10⁻¹³	-0.0143
0.175	2.9 x 10⁻⁹	-0.00376
0.167	2.0 x 10⁻⁹	-0.0045
0.188	3.5 x 10⁻⁹	-0.00674
( ) 0.197	6.8 x 10⁻¹⁰	-0.0028
( ) 0. 24		-0.0072

(1) Same film. Values given are for near room temperature (first value)
and at low temperatures (second value)

( ) Same film. Values given are for near room temperature (first value)
and at low temperatures (second value)

The most significant feature of this data, however, results from the fact that if carrier generation is determined by a "Poole-Frenkel" type barrier lowering in the bulk of the film, then, it would be $\frac{0.0076b}{K^{1/2}}$ eV/v$^{1/2}$, where b is defined by Eq. 134. Taking K to have the low-frequency value of fused silica, 3.9, and with b >> 1, values of $dE_a/dv^{1/2}$ were observed which were too low to be explained. If, however, it is assumed that the barrier lowering is due to the "image" force at the electrodes seen on end, then $dE_a/dv^{1/2}$ would have the value $\frac{0.0023b}{K^{1/2}} \frac{eV}{v^{1/2}}$ which is much more consistent with some of the measured values. This indicates the electrodes play a greater role in the conduction process than has hitherto been indicated.

As discussed in Section V, the potential across the substrate surface if there were no current would be an Arccosine function, resulting in a barrier lowering proportional to the two-thirds power of the applied voltage in the neighborhood of the electrodes. Such a two-thirds power dependence has not been observed, implying either that the carrier generation occurs near the center of the film where the field is nearly constant, or the field is determined by charge buildup in the film. The former contradicts the previous paragraph. In the latter case the electric field is proportional to the film sheet resistance. This can be seen from the fact that

$$\overline{\mathcal{E}} = \rho \overline{J}$$

where $\overline{J}$ is the current density, ρ is the resistance per square, and is the electric field. Since in steady-state $\nabla \cdot \overline{J} = 0$ and the current flow is one-dimensional, $\overline{J}$ is a constant and, hence, $\mathcal{E}$ is proportional to ρ.

If the field is indeed determined by the sheet resistance, then normally the film would be uniform between the electrodes, hence, $\mathcal{E}$ would be constant. From Eq. (114), the constant b would be unity. As will be seen later in the discussion of the effects of high electric fields and annealing on the electrical properties of these films, there is evidence for considerable surface mobility of the islands even at room temperature. Normally such island mobility would not disturb the film uniformity since as many atoms would diffuse on the average across a line drawn on the film in one direction as in the opposite direction. If the line were drawn just outside an electrode, parallel to the electrode, however, there would be no out-diffusion from the electrode, while any diffusion from the film to the electrode would adhere on the electrode. This material transport would result in a depletion of the region adjacent to the electrodes, increasing the resistance and the electric field in this region. Since an increase in the electric field would increase the constant b by an undetermined amount, this increase is consistent with the measurements shown in Table I.

81

After a film is deposited at room temperature the resistance does
not remain constant if the substrate remains at room temperature.
Figure 22 follows a typical film resistance decay at room temperature
over more than three decades of current. As can be seen, the logarithm
of the current is proportional to the square root of time. Since the
vacuum during this period was on the order of 10^{-11} torr, it is not
likely that the residual atmosphere was responsible for the current
decrease. The square root of time dependence is suggestive of a dif-
fusion process involving either the diffusion of the deposited material
into the substrate or the diffusion of adsorbed deposit molecules and
islands into the film. The low diffusion coefficient of gold in fused
silica at room temperature would tend to eliminate the former process.

While the current is decreasing the activation energy increases.
The amount of increase varies from a negligible amount to the situation
depicted in Fig. 15. Since the amount of deposited material can be con-
sidered constant during the room temperature anneal, the normal assump-
tion is that the process of nuclei agglomeration continues, resulting
in an average increase in the island size with an increase in the aver-
age inter-island spacing. The increase in island size would be expected
to decrease the activation energy, provided the longest island dimension
increases. A possible explanation for the observed behavior would lie
again in the assumption of a materially depleted layer immediately out-
side of the electrodes. Annealing would be expected to extend this
depleted layer further into the film. This would distribu the applied
field over a longer distance, decreasing ß. This would decrease the
field lowering of the potential barrier, and, hence, increase the acti-
vation energy. The increase in the inter-island spacing would also
decrease the screening effect of neighboring islands, increasing the
activation energy. An increase in the inter-island spacing would de-
crease the particle transport function and thus decrease the current
beyond that due to the activation energy alone. The amount of change
in the activation energy seems to follow the $dE_a/dv^{\frac{1}{2}}$ for those films
where long-term annealing was investigated.

The effects of room-temperature annealing are speeded up by anne-
ing at higher temperatures. This is shown in Fig. 23. An irreversible
increase in the film resistance is observed together with an increase
in activation energy which may be slight if the $dE_a/dv^{\frac{1}{2}}$ of the film is
small. For the film in Fig. 23 the increase in E_a is negligible, while
$dE_a/dv^{\frac{1}{2}}$ is about 0.002 ev/v . The very slight increase in resistance
with four days of room temperature annealing can also be seen. High
temperature anneals, in general, reduce the effect of further room
temperature anneals, as might be expected.

A surprising effect which was observed was the influence of high
fields on the films. This can be seen in Fig. 24, which shows the time
effect of a 600-volt constant applied voltage at constant temperature.
Subsequent removal of the high voltage leaves the film at a lower resis-
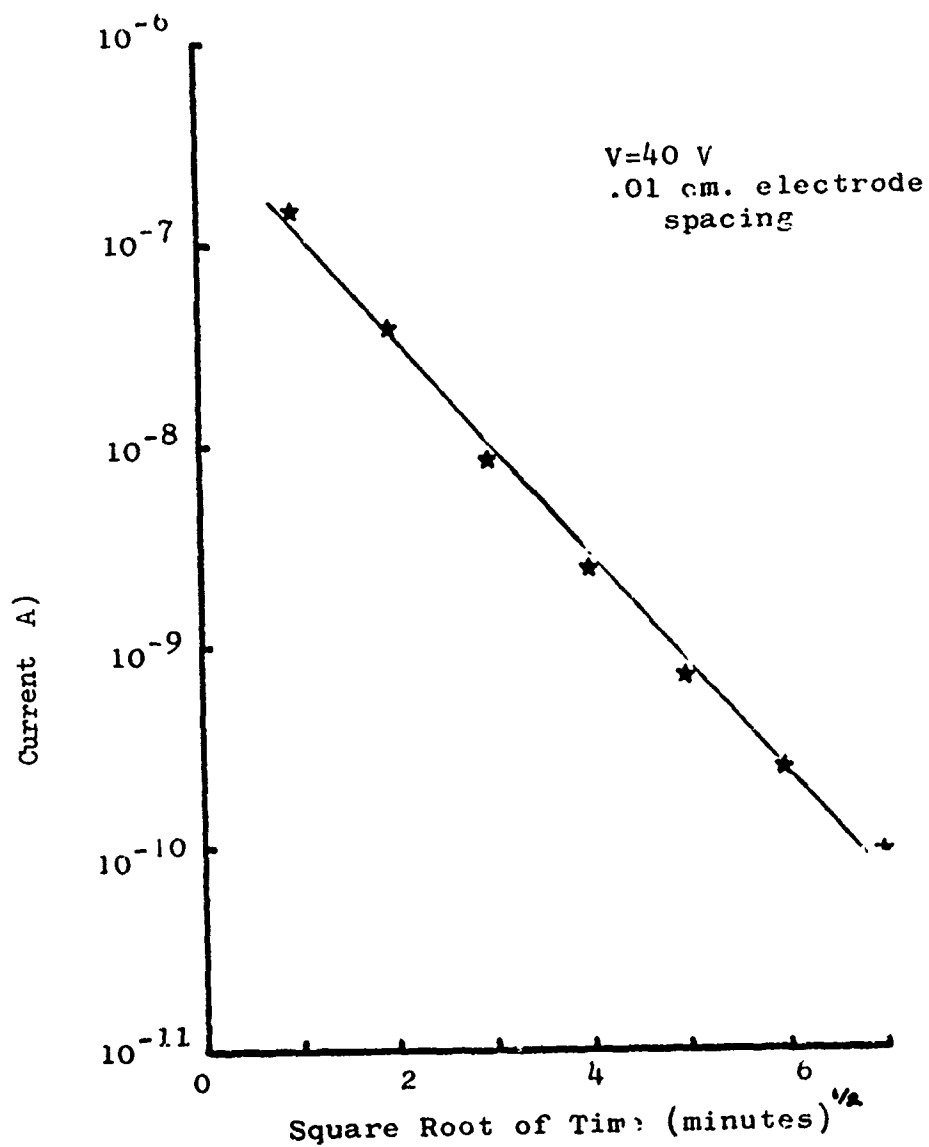tance state at low voltages than it was in prior to the application of

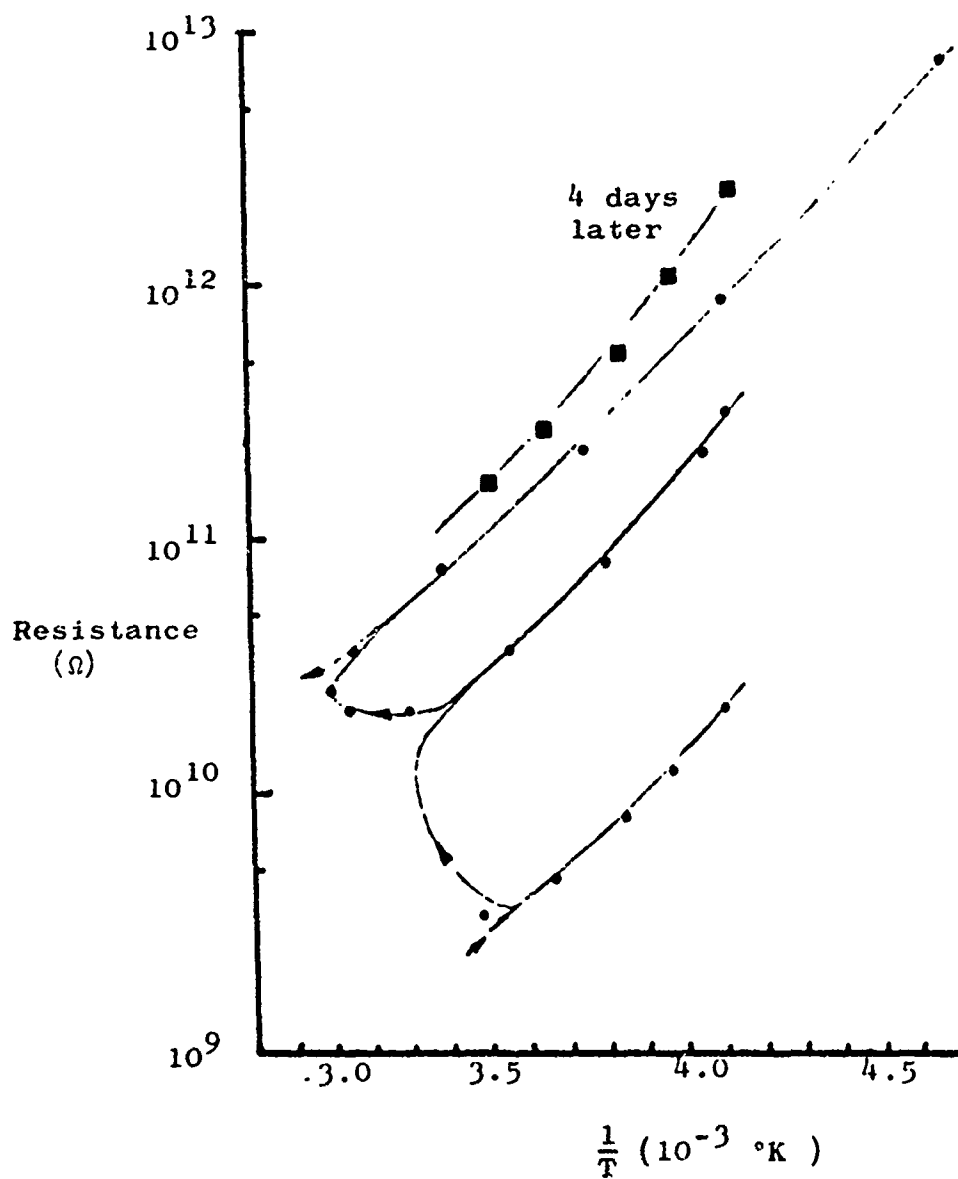Fig. 22 - Room Temperature Anneal of Film
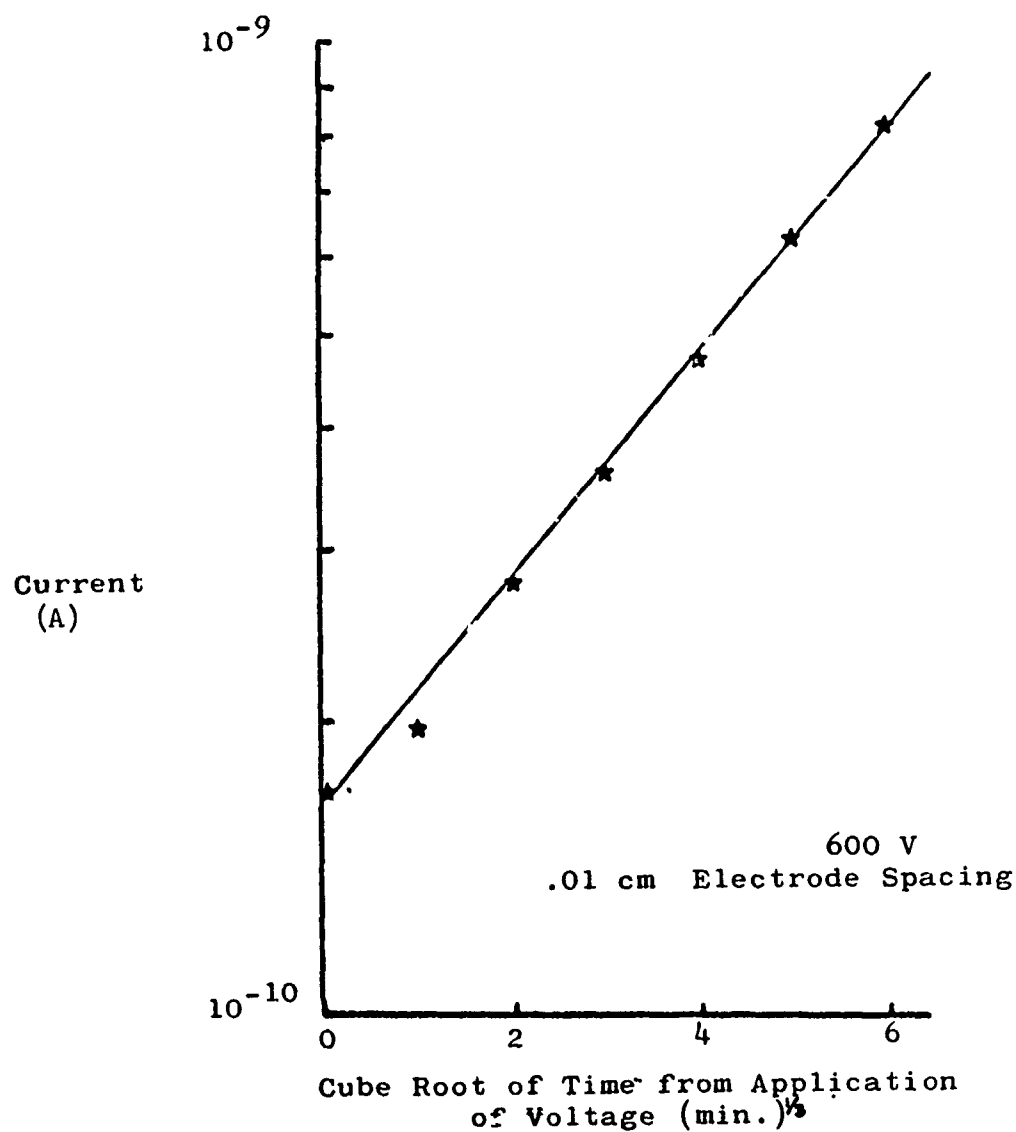
Fig. 23 - Anneals at Higher Than Room Temperature

Fig. 24 - Film Current Increase With Time Due to Large Applied
Voltage

the high voltage. The film then anneals as if it were a freshly deposited film at the lower resistance. High voltages tend to reverse the effect of anneals.

The mechanism for this high-field effect is questionable. The increased current and field induced lowering of the potential barrier must be associated with an increased number of free carriers in the film and, hence, with an increased number of electrostatically charged islands. The electrostatic charges would result in electrostatic forces which tend to break up islands. If the potential were determined by a depleted region in the vicinity of the electrodes, the high applied field would result in a very high electric field at the edge of the electrode, as given by Eq. (121). Since this results in a pressure on the metal of $\frac{1}{2}\epsilon\xi^2$, high fields can cause points on the electrodes where the yield point of the metal is exceeded, and islands can be pulled from the electrodes, in contrast to the annealing process.

Another contributing factor would be the change in the island shape with time. If material is added to the edges of a film by surface diffusion of adatoms during the film deposition, then over a period of time diffusion of atoms from the edges of the islands to the top of the island would be expected since this decreases the surface-to-volume ratio of the island and, therefore, decreases the positive surface tension energy term. Since it has been shown that a minimum electrostatic energy configuration is a distended, flattened disc, the in-diffusion of the edges would increase the electrostatic energy, as has been experimentally observed. An increased voltage would cause an increased number of charged islands due to the field-induced barrier lowering. The presence of charge on an island would tend to push the edges of the island out again, reversing the effect of the anneal, as is experimentally observed.

A further effect of very high fields observed in some films was the collapse of film resistance so that the film exhibited a high conductance state which then persisted at lower fields. This previously unreported phenomena was investigated extensively in two films, one of very high sheet resistance, the other of a relatively low sheet resistance.

The first film was grown on a Pyrex substrate. After the first deposition on this substrate the current showed less than 5×10^{-14} amps up to 150 volts. Such low currents were close to the limit of resolution for the Kiethley 610 B electrometer used in the current measurements. Between 150 volts and 310 volts, the current was unsteady and prone to rapid fluctuations. At 310 volts the current steadied to about 7.8×10^{-1} amps. From this point the film responded like a normal film of lower resistance, even when the applied voltage was lowered to 9.5 volts; this behavior is shown in Fig. 25. The hysteresis in the low resistance values probably result from a high voltage film conditioning. When the film was short circuited for a period of time, it reverted to the high
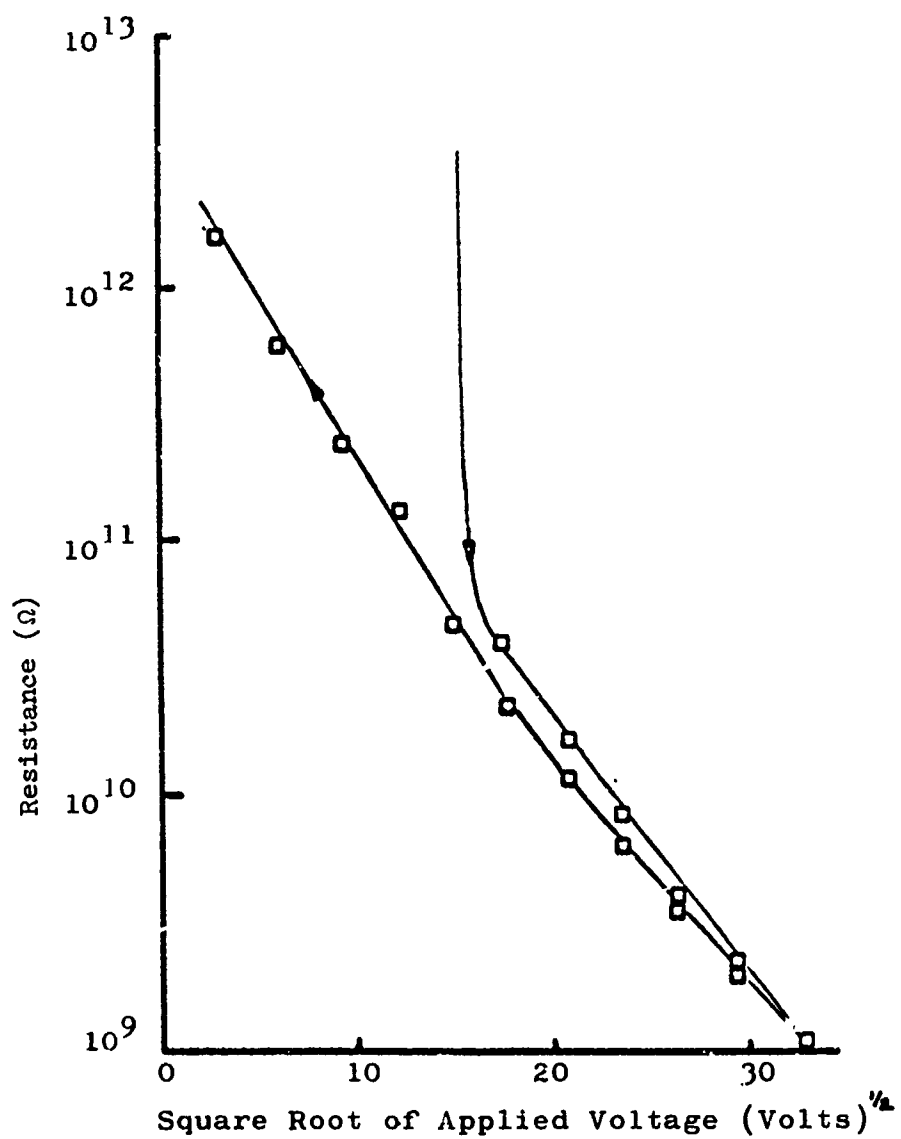
Fig. 25 - Film Resistance Subsequent to Breakdown

resistance state again for low voltages. The film was cycled in and out of the low resistance state several times, with the applied voltage necessary to switch the film variable. During temperature cycling it was found that lower voltages were required at low temperatures, although the voltage at breakdown was never as low as in the initial breakdown. The film broke down between 700 and 850 volts at room temperature, at between 550 and 700 volts at about -10°C and at between 225 and 300 volts at -50°C. Once in the low resistance state the film could not be easily converted to the high resistance state. Occasionally, removing the applied voltage would accomplish this conversion, but at other times the film would show the low resistance after having the excitation removed for several hours. Sometimes reversing the polarity would put the film in the high resistance state; on other occasions the low resistance would be symmetrical with positive and negative applied voltages. When a negative voltage was applied to the film on one occasion the resistance was symmetrical, and then suddenly switched to the high resistance state while the voltage was still on the film.

Another film was grown and annealed for about 21 hours, with the current showing the decay illustrated in Fig. 26. The rapid initial decay of current before the onset of the linear log I vs. the square root of time is a common transitory phenomenon. After the annealing period the voltage across the film was increased to about 300 volts, at which point the film displayed sudden, drastic decrease in resistance of the form discussed previously. When the film was subsequently subjected to a 40-volt impressed potential, the same potential at which the data in Fig. 26 was taken, the film would spontaneously switch in and out of the low resistance state. This behavior is shown in Fig. 27, which was taken from a strip chart recording of the ammeter output. Subsequent to this measurement the film was removed from the vacuum and examined in a metalograph. This film is shown in Fig. 28, which shows a considerable disruption of the film. There are several noteworthy points of interest regarding this film. The breakdown of the film caused a decrease in the resistance of the film of over four orders of magnitude above that immediately prior to breakdown. The "on" resistance of the film is approximately 22,000 ohms, which is readily detectable without sophisticated instrumentation. These two factors raise the possibility that such films could be used as programmable memory elements if the phenomena could be satisfactorily controlled. In addition, it can be seen in Fig. 28 that the film shows evidence of breakdown in a continuous line outside the electrodes, even if that breakdown did not extend all the way across the interelectrode spacing. It is possible, considering the small geometries involved, the relatively poor thermal conductivity of fused silica, and the relatively large currents and voltages involved, that local "hot spots" were developed by resistive heating at the points of current passage through high fields, and the "hot spots" caused the disruption depicted. If this were the case, the field must have been highest at the edges of the film, as was proposed earlier, and that assuming current were more easily injected across the disrupted area, so that the "injection" resistance appeared to be lower, the resistance across the center of the film must have been negligible in comparison
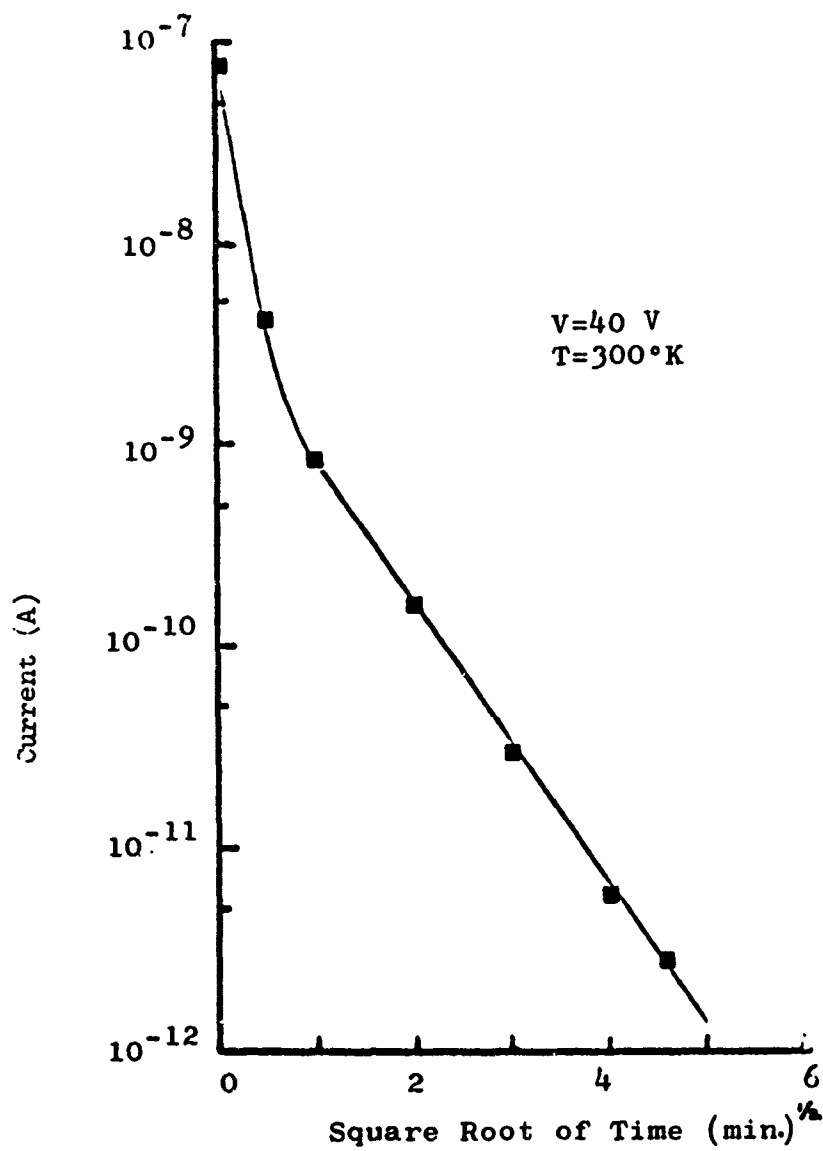
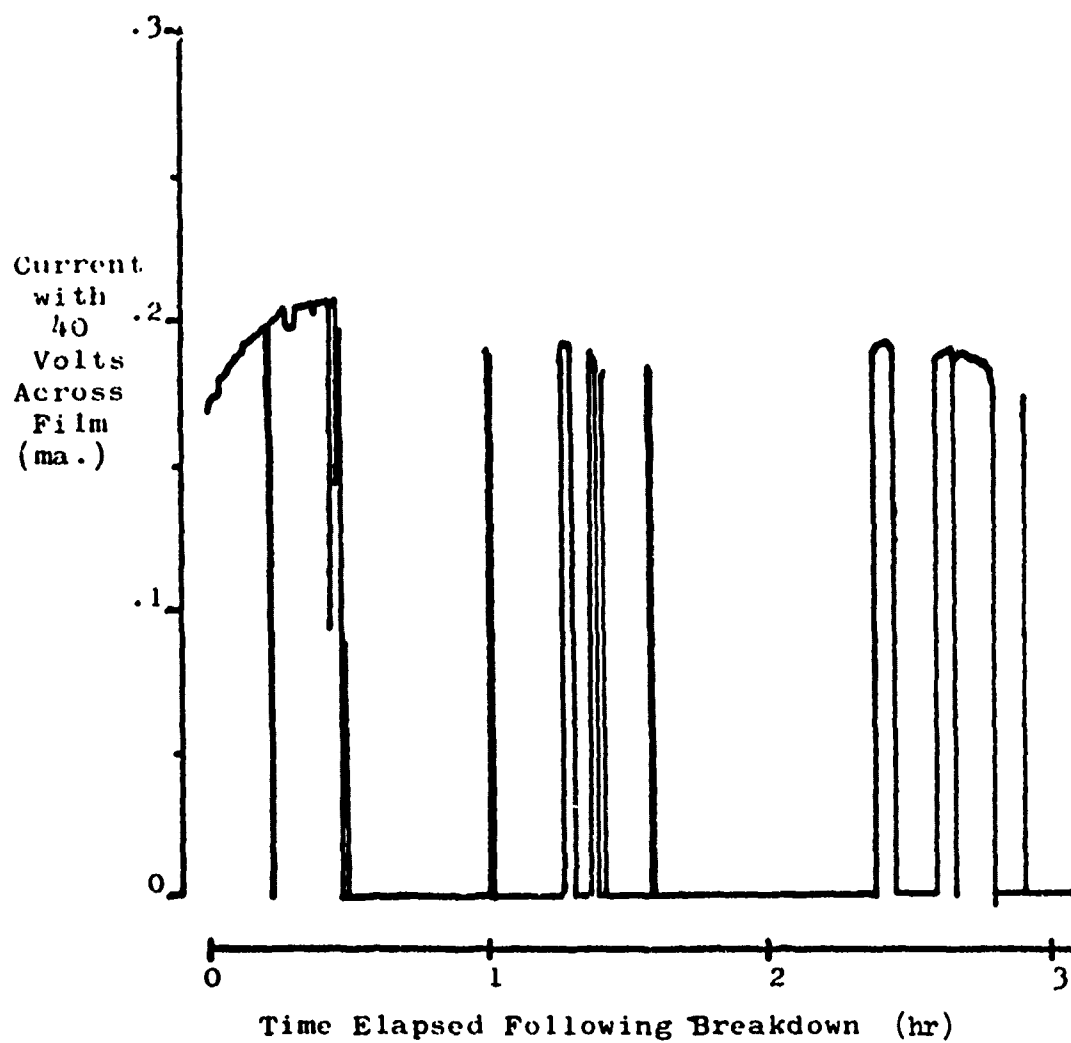Fig. 26 - Effect on Current of Room Temperature Anneal

Fig. 27 - Spontaneous Switching of a Film Following Breakdown

90

to the regions "shorted out" by the disrupted area. If the regions where the disruption shorted from one electrode to the other were the sole cause for the lower observed resistance, then there would have been no reason for "hot spots" to develop in adjacent regions which show disruptions only along the electrode edges. All other explanations evolved for such a film structure have also included an enhanced field in the regions where the breakdown occurred; hence, a higher field at the electrodes than in the center of the film. It should also be noted that the disrupted areas shown in Fig. 28 did not appear in those regions of the electrodes which were masked from the deposit, even though there were regions which were otherwise identical to the deposited area in such features as spacing between the electrodes, electrode thickness, electrode sturcture, etc. The breakdown illustrated can, therefore, be attributed only to the film presence.

Some interesting preliminary work has been done on the influence of the electrode separation on the film properties. A substrate was made which had three electrodes, separated into a center electrode, an electrode spaced 0.01 cm from the center electrode, and an electrode spaced 0.025 cm from the center electrode. The lengths of both gaps were essentially equal. Electrical contact was made to all three electrodes, and a discontinuous film deposited over the area containing the three electrodes on the assumption that conduction between the electrodes other than across the narrow gap would be negligible.

The results of this investigation are shown in Table II. Initially, negligible current was measured across either gap. Immediately after the film was deposited the 0.025-cm gap film showed a lower resistance than the 0.01-cm gap film for small voltages, contrary to what would normally be expected for the same voltage applied across both films. With increasing voltages, the 0.01-cm film exhibited a larger percentage decrease in resistance than the C.025-cm film, so that above about 28 volts the 0.01-cm film showed a lower resistance than the 0.025 cm film. As the films annealed at room temperature, the current through the films decreased, as would be expected; however, the current through the 0.025-cm film decreased at a faster percentage rate than that of the 0.01-cm film, so that after 17 hours the current through the 0.025-cm film was immeasurably small over the entire voltage range, while the current through the 0.01 cm film was still measurable over part of the voltage range. The same general result was found in a second film. After the measurements shown in Table II were taken, additional material was evaporated on top of this film and the resistance measured. It was found that both films then exhibited the "breakdown" phenomena, with the 0.025-cm film "breaking down" at a lower voltage than was required for the 0.01-cm film.

These results are difficult to explain if it is assumed that the sheet resistar   of the deposited films is constant. If, however, the deposited film   depleted in the area adjacent to the electrodes, and this depletion extends several microns into the film, then it would be possible to have an overlapping of the depleted regions in the 0.01 cm
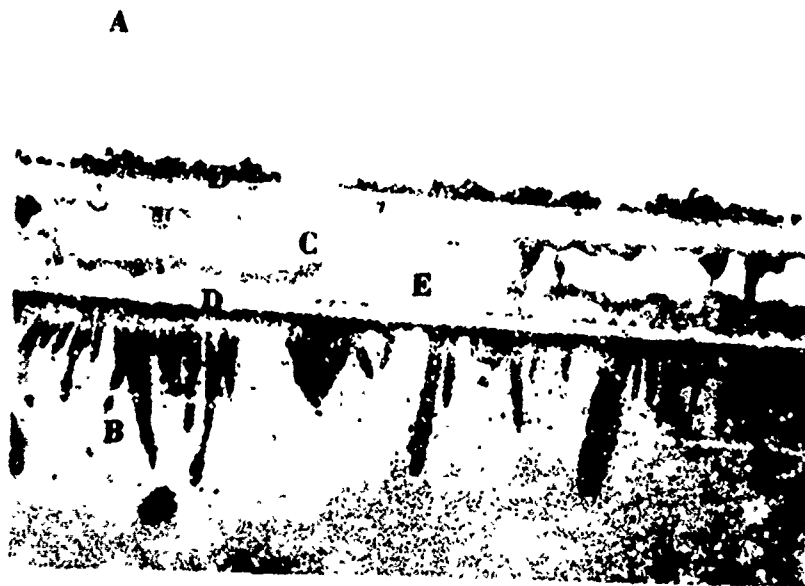
A. Upper Electrode
B. Lower Electrode
C. Discontinuous Film
D. Overlap of Electrode Au and Cr Layers
E. Thermal Disruption of Film

Fig. 28 - Film Following Breakdown

92

Table II - Current Through Films With 0.01 cm and 0.025 cm Electrode Spacing

V Setting	0.025 cm Current (amp)	0.01 cm Current (amp)
Immediately After Deposition		
1 (0.351)	9.2×10^{-13}	4×10^{-14}
2 (1.15)	3.3×10^{-12}	1.5×10^{-13}
3 (2.15)	6.1×10^{-12}	3.8×10^{-13}
4 (4.37)	1.25×10^{-11}	1.3×10^{-12}
5 (7.12)	2.14×10^{-11}	3.4×10^{-12}
6 (12.0)	3.9×10^{-11}	1.12×10^{-11}
7 (17.6)	5.9×10^{-11}	3.1×10^{-11}
8 (23.25)	8.4×10^{-11}	6.6×10^{-11}
9 (28.9)	1.12×10^{-10}	1.25×10^{-10}
10 (40.0)	1.8×10^{-10}	3.5×10^{-10}
$1\text{-}\frac{1}{2}$ Hours Later		
1	≈ 0	≈ 0
2	≈ 0	1.8×10^{-14}
3	≈ 0	6.5×10^{-14}
4	≈ 0	2.4×10^{-13}
5	≈ 0	6.7×10^{-13}
6	≈ 0	2.34×10^{-12}
7	1.4×10^{-12}	6.6×10^{-12}
8	2.4×10^{-12}	1.47×10^{-11}
9	3.7×10^{-12}	2.9×10^{-11}
10	6.8×10^{-12}	8.6×10^{-11}
17 Hours Later		
1	≈ 0	≈ 0
2	≈ 0	≈ 0
3	≈ 0	≈ 0
4	≈ 0	≈ 0
5	≈ 0	≈ 0
6	≈ 0	3.8×10^{-14}
7	≈ 0	1.65×10^{-13}
8	≈ 0	4.2×10^{-13}
9	≈ 0	8.8×10^{-13}
10	≈ 0	2.9×10^{-12}

film with a considerable increase in the sheet resistance at any point outside an electrode over that which would be seen at the same distance outside an electrode with a wider gap. The effects of a continued diffusion of material to the electrodes would also be expected to be different for the two cases.

SECTION VIII

CONCLUSIONS

In this work electrical conduction in discontinuous metal films on an insulating substrate was investigated. An analysis was made of the electrostatic energy associated with an electronic charge on a metal island, and the interaction between a charge and an island. This interaction was emphasized because it was assumed that there are two problems involved in charge transport; the creation of a positive charge and the transfer of that charge from one island to the next. The barrier to localization of a charge on an island is the energy required to separate a charge from an initially uncharged island to create two electrostatically charged islands if the carrier is created in the film. If the carrier is injected at an electrode, however, the energy is basically the electrostatic energy of one charged island, which is about one-half that for carrier creation in the film. The number of charge carriers was calculated using Maxwell-Boltzmann statistics to give a function that is approximately exponentially proportional to the electrostatic energy divided by Boltzmann's constant times the absolute temperature. The resistance is known to be nonlinear with the logarithm of the resistance often observed to decrease proportional to the square root of the applied voltage. When the charge carriers are created in the center of the film, the nonlinear resistance is attributed to a field lowering of the electrostatic potential, as in the Poole-Frenkel effect. It was shown that the potential must have a $1/r$ dependence outside an island to give a barrier lowering proportional to the square root of the voltage. For high fields and nonspherical islands, or considering the effect of islands intervening between the charge centers, the potential would not have the necessary $V^{1/2}$ dependence, and, hence, should lead to deviations from the observed behavior. If, on the other hand, the carriers are injected from the electrodes, then the nonlinear resistance can be attributed to a modified Schottky barrier lowering at the electrode. This barrier lowering can be expressed by

$$\Delta E = 2.33 \times 10^{-4} \frac{\mathcal{E}^{1/2}}{K^{1/2}} \frac{(eV\ cm)^{1/2}}{V^{1/2}} \quad .$$

The enhanced electric field which can be expected in the vicinity of the electrodes would decrease the effect of intervening islands. It is necessary, however, to assume a linear field in the vicinity of the electrodes.

The question arises as to how the charge gets from one island to the next. The mode of transport affects the calculation of the energy barrier. If tunneling is assumed, the electrostatic energy must be calculated using the high-frequency dielectric constant because of the

95

short transit time. This calculation would result in the electrostatic energy being about 3.9 times larger, assuming the use of the bulk dielectric constant of quartz. Thermal emission would require either some allowed states within the forbidden gap of the substrate, or else the contact potential or the work function would have to be surmounted. This last situation would require a higher activation energy than has been often observed. Arguments were presented to show that there is a high concentration of trapping states in the substrate, and it would be possible for a carrier to "hop" from one site to the next. The carrier need only surmount the electrostatic fields from the islands. If the "hopping" distance were short, the low frequency dielectric constant could then be used.

Once a charge were localized on an island, the barrier to the transfer of that charge to a neighboring island would be due to the polarization of an island caused by a charge outside the island. This barrier is much less than that for the initial creation of the charge since the image/force is short-ranged, decreasing as $1/D^5$ where D is the distance from an island. Assuming the same "hopping" transport considered above, a charge would be relatively mobile once it was initially localized on an island. If, on the other hand, tunneling between islands were the charge transport mechanism, then the electrostatic energy involved in the polarization of the substrate, as represented by the low frequency dielectric constant used in the activation energy, is required every time the charge moves from one island to an adjacent island. If the bulk dielectric constant is used, a major portion of the activation energy must then be supplied for every inter-island transfer, and the charge is relatively immobile. Supplying the energy thermally for each tunneling "hop" will not give the required $V^{\frac{1}{2}}$ dependence of the logarithm of the resistance.

Because of the uncertainty in the value of the dielectric constant and the distribution in the island sizes, the proper form of the electrostatic energy could not be experimentally determined. It was shown, however, that the field lowering of the potential barrier is most consistent with electrode emission.

The observed behavior of the voltage dependence of films simultaneously deposited across adjacent gaps of different lengths is difficult to reconcile with an interpretation other than that the conductivity is determined by an electrode effect which extends a finite thickness into the films. In this experiment the films as first deposited showed that, for the same applied voltage across a 0.01 cm gap and a 0.025 cm gap, the 0.01 cm film initially showed a higher resistance, even though it had the higher field, which normally implies a lower resistance, even though the sheet resistances should have been the same in the deposits between both gaps. Room temperature annealing caused the larger gap to eventually show the larger resistance.

A significant experimental result involved the breakdown of some of the films when a sufficiently high electric field was applied across the

film. An analysis of these films suggest that the high fields overcome the electrode barriers by field emission or some other reversible electrode effect, so that charges may be easily injected into the film islands. Examinations of the films which underwent breakdown showed evidence of most film disruption, and, hence, highest fields, along the electrodes. This is again evidence that the electrode effects dominate the film conduction. The phenomena of film breakdown is not completely understood, and would bear further investigation for both its theoretical importance and its possible technological application.

It was pointed out in the theoretical sections that if charge were localized on an island, the surface of the island would experience an outward stress. The magnitude of the stress would depend on the geometry which is assumed, but it was pointed out that in many cases the stress would approach the yield point of the metal. This charge, especially in the case of ellipsoids, would tend to elongate the ellipsoid along its major semiaxis with a force inversely proportional to the island's radius of curvature. Localization of the charge on an island would also cause attractive forces on the surface of an adjacent, uncharged island. These forces would have an effect on the nucleation and agglomeration stages of thin film growth which have not been previously discussed.

The presence of a finite conductivity in the discontinuous films even at low applied voltages supports the hypothesis that a number of charged islands exponentially proportional to the activation energy is always present, even in the complete absence of an applied field. Such charges on an island will oppose the surface tension force tending to pull the island into a sphere with the smallest surface-to-volume ratio. The charges tend to elongate the island, and to draw adjacent islands into contact with the charged island. For this reason the island will be more successful in capturing sub-critical nuclei. This tendency to elongation may also provide a theoretical explanation for the "bridges" which have been claimed by some authors[79] to join islands prior to the merging of two islands. Such extensions from an island have only been observed toward an adjacent island, and would require an increase in the surface tension energy of the metal island. Both these facts tend to verify the electrostatic explanation for this phenomena. The tendency toward more filamentary island growth when the number of charged islands increases either by the lowering of the activation energy as the islands become larger, by an impressed voltage, or by the introduction of external charges by electron bombardment is also consistent with this theory.

Mention should also be made that[12] in situ electron microscopic studies of thin film nucleation have shown a surprising island mobility, which has been termed "liquid-like" motion. The ionizing effect of the electron beam used for the microscopy would result in charge concentrations far in excess of what would normally be encountered. This charge would accentuate the electrostatic forces and contribute to the motion that has been observed.

It was noted that annealing the films at room temperature for a period of up to several days caused an increase in the resistance and an increase in the activation energy of the films. An interpretation for this could be the diffusion of the metal from the island edges, where it was first attached during the island growth, to the center of the island, increasing the metal thickness at the center. This diffusion would result in a lower surface tension energy term since the surface-to-volume ratio would decrease. At the same time the distance between islands would increase, lowering the "hopping" probability, and the largest island dimension would decrease, increasing the electrostatic activation energy. Higher activation energy and lower currents are consistent with the experiments. The effect of charge on the islands can be seen by the fact that high current densities, which would imply more charged islands, reverse the effects of the annealing even if the film is measured at the lowest possible voltage subsequent to the application of the large impressed voltage. This high current effect would be expected if the charged islands spread out as a consequence of the electrostatic induced stresses. The experimental results can also be explained by diffusion of the deposited film into the electrodes. High fields could be considered to pull material from the electrodes by the field-induced stresses on the electrodes. In a film both processes probably participate.

In summary, a theoretical re-evaluation of the electrostatic energies involved in current transport through discontinuous films was carried out. The results of this analysis were combined in a theory of the mechanism for the current flow in discontinuous films whose salient features were the elimination of any long-distance tunneling transport and the inclusion of electrode injection effects. Experimental evidence was presented for films which gave evidence of a much larger electrode influence than had previously been claimed. These films showed several interesting properties, the most remarkable of which was the reversible change of these films from a high to a low resistance state with a sufficiently high electric field, and the ability to subsequently maintain that low resistance to very low, even zero, voltages. The implications of the electrostatic energy calculations on film nucleation and agglomeration theory were also mentioned, since the electrostatic forces could provide a driving force for island distortions.

Experimental investigations are continuing to confirm many of the early observations. In particular, attempts are being made to carry out the depositions on sodium chloride substrates, since the film stripping for electron microscopy would be greatly facilitated, and it would be possible to mount and investigate the films between the electrodes. This experiment would directly confirm the influence of the electrodes on the film structure adjacent to the electrodes. The microscopic effects of the breakdown phenomena, if it persists in the salt substrates, could also be directly investigated.

Further work will also be done on the breakdown phenomena in the films described in this report. It is hoped that the mechanism for the breakdown can be isolated, and that a knowledge of the cause will lead to a better control of the phenomena with a possible application of this phenomena in useful devices.

It is hoped that the thin film measurements can be refined to the point where it would be possible to measure such film properties as island size and separation by the use of the calculations presented here so that the film resistance properties can be used for the monitoring of nucleation for a better understanding of nucleation theory.

APPENDIX I

THERMODYNAMICS OF NUCLEATION

It is a common feature of many very thin films and thin continuous films which have been sufficiently heated that these films are in the form of unconnected "islands" or nuclei. The process of nucleation, or island formation, has received extensive study, both theoretically and experimentally. Reference 8 contains an extensive bibliography on the various approaches to this problem.

One analysis performed by Hirth and Moazed[7] illustrates the underlying cause for the island formation. The Gibbs free energy for an island can be expressed as the sum of two terms, one negative and proportional to the cube of the island radius, the other positive and proportional to the square of the island radius. The former term represents the condensation energy of the metal vapor while the latter term represents the surface tension contribution to the total energy. For large islands the condensation energy term would dominate and the Gibbs free energy would decrease with increasing island radius. For very small islands, the large surface-to-volume ratio causes the surface tension term to dominate, causing the free energy to decrease with decreasing island radius. The natural tendency for processes to continue in the direction of reduced free energy is responsible for the observed increased vapor pressure of sufficiently minute particles causing an enhanced re-evaporation. The derivative of the Gibbs free energy with respect to island radius can be set equal to zero to estimate the minimum island radius, called the critical nuclei size, which will be stable in the respect that it will not tend to spontaneously dissociate. For a film, the total free energy will be the sum of the free energies for the individual islands plus an entropy of mixing term. This expression for the total free energy of the system allows the calculation of rate of formation of critical nuclei. Above the critical nuclei size, the islands will tend to grow by coalescence of the islands or adsorption of monomers diffusing over the surface. Island growth is limited by depletion of the condensed material in a zone around the nuclei as the island grows, and perhaps by an enhanced pressure tending to split the islands because of electrostatic charges on the islands. Many other approaches to nucleation can be shown to be fundamentally the same as that just discussed,[8,9] differing in attempts to account for the atomic dimensions of the nuclei, thermal accommodation, and other substrate effects.

Experimentally[10,11] it has been noted that the observed nuclei size distribution is more uniform than would be expected on the basis of nucleation theory. This is especially true if the very high surface diffusion mobilities implied in situ electron microscope studies of nucleation[12] are actually present. Island movements have been so mobile that the term "liquid like" has been applied to this behavior. It should be noted that, at one substrate temperature, a deposit may form a continuous

film, while at higher temperatures that same film may spontaneously break up into a discontinuous, island structure.[13] These islands are also of a remarkably uniform size even though thermodynamically the lowest energy would exist for the lowest surface-to-volume ratio, and, hence, the largest possible island size.

The electrostatic influences on nucleation are not normally included in analysis since it is assumed that free charges are not present in a film without an impressed current. A strong effect was experimentally found for nucleation in an external electric field[14] and for nucleation while the substrate underwent electron bombardment.[15,16,17] It was found in these studies that the electron bombardment resulted in more numerous, smaller islands which became continuous at a smaller average thickness. Dove[18] has proposed that the image force of charge on neighing islands can be a driving force for agglomeration. The experimental evidence from the electrical effects on nucleation suggest that agglomeration can also be opposed by charges. The more detailed investigation of this problem is one topic of this investigation.

INCREASE IN THE WORK FUNCTION OF AN ISOLATED SPHERE

To calculate the effects of a charge, for an insulated sphere of radius, a, influenced by a charge, e, a distance, R, from the center of the sphere, the surface of the sphere can be made free of tangential fields by establishing an image charge Q' of magnitude $-\frac{a}{R}$ e at a distance d = a^2/R from the center of the sphere along a line toward the external charge. With a total charge Q on the sphere, an additional charge of magnitude Q-Q' must be at the center of the sphere to raise the potential of the sphere to its proper value. Therefore, the force on the electron due to the image charge and the charge at the center of the sphere is

$$F(R) = \frac{a^3 e}{4\pi\epsilon}\left[\frac{2R^2 - a^2}{R^2(R^2 - a^2)^2}\right] + \frac{eQ}{4\pi\epsilon R^2} \quad . \tag{A1}$$

When calculating the work required for an electron to excape from the island, defined as the work function, the island will be left with a net charge +e, and the force must be integrated to R = ∞ . It will be noted, however, that the integral becomes infinite if the lower limit is the surface of the sphere, R = a, because, as mentioned earlier, the image force is only the long range limit of the correlation force. The lower limit is assumed for the present to be a + ξ, so that the work function is given by

$$\varphi = \frac{e^2}{4\pi\epsilon}\int_{a+\xi}^{\infty}\left\{a^3\left[\frac{2R^2 - a^2}{R^3(R^2 - a^2)^2}\right] + \frac{1}{R^2}\right\}dR \quad . \tag{A2}$$

When "a" becomes infinite, and R is replaced by a + ξ the resulting case is that of an electron escaping from a planar surface, and the solution of Eq. (A2) yields the bulk work function φ_0 expressed in terms of ξ , or conversely

$$\xi = \frac{e^2}{16\pi\epsilon\varphi_0} \quad . \tag{A3}$$

Substituting Eq. (A3) in Eq. (A2) yields

$$\varphi = \varphi_0 + \frac{e}{4\pi\epsilon}\left[\frac{e}{a} - \frac{5}{8}\frac{e}{a}\right] = \varphi_0 + \frac{3}{8}\frac{e^2}{4\pi\epsilon a} \quad . \tag{A4}$$

The first term inside the bracket has been identified with the electrostatic energy in the creation of two charged islands, while the second

term inside the brackets can be attributed to the fact that while an electron outside a planar conductor is always attracted to the conductor by an image charge, an electron outside a sphere is attracted to the sphere by only a dipole, which has a much shorter range.

REFERENCES

1. Garron, R., Ann. Phys., vol. 10, 595-621 (1965).

2. Offret, M. and Vodar, M. D., J. Phys. Radium, vol. 17, 237 (1955).

3. Borzjak, P. G., Sarbej, O. G., and Fedorowitsch, R. D., Phys. Status Solidi, vol. 8, 55-58 (1965).

4. Neugebauer, C. A., Physics of Thin Films, vol. 2, Academic Press, 1964.

5. Sennett, R. S. and Scott, G. D., J. Opt. Soc. Amer., vol. 40, 203-211 (1950).

6. Maissel, L. I. and Gland, R., editors, Handbook of Thin Film Technology, McGraw-Hill, Inc., 1970, Chapter 8.

7. Hirth, J. P. and Moazed, K. L., Fundamental Phenomena in the Materials Sciences, vol. 3, Surface Phenomena, Plenum Press, 1966.

8. Sigsbee, R. A. and Pound, G. M., Advan. Coll. Interf. Sci., vol. 1, 335 (1967).

9. Lewis, B., Thin Solid Films, vol. 1, 85 (1967).

10. Poppa, H. J. Vac. Sci. Tech., vol. 2, 42 (1965).

11. Herman, D. S. and Rhodin, T. N., J. Applied Phys., vol. 37, 1594-1602 (1966).

12. Pashley, D. W., Stowell, D. W., Jacobs, M. H. and Law, T. J., Phil. Mag., vol. 10, 127 (1964).

13. Kane, W. M., Spratt, J. P., and Hershinger, L. W., J. Applied Phys., vol. 36, 2085 (1966).

14. Chopra, K. L., Appl. Phys. Letters, vol. 7, 140 (1965).

15. Lewis, B. and Campbell, F., J. Vac. Sci. Tech., vol. 4, 209 (1962).

16. Hill, R. M. Nature, vol. 210, 512 (1966).

17. Stirland, D. J., Appl. Phys. Lett., vol. 8, 326 (1966).

18. Dove, D. B., J. Applied Phys., vol. 35, 2785 (1964).

19. Neugebauer, C. A. and Webb, M. B., J. Appl. Phys., vol. 33, 74 (1962).

20. Neugebauer, C. A., Physics of Thin Films, vol. 2, G. Haas and R. Thum, ed., Academic Press, 1964.

21. Neugebauer, C. A. and Wilson, R. H., Basic Problems in Thin Film Physics, Vandenhoeck & Ruprecht in Goyyingen, 1966.

22. Hartman, T. W., J. Appl. Phys., vol. 34, 943 (1963).

23. Herman, D. S. and Rhodin, T. N., J. Applied Phys., vol. 37, 1594 (1966).

24. Wei, L. Y., J. Chem. Phys., vol. 39, 2709 (1963).

25. Minn, S. S., J. Rech. Centre Natl. Rech. Sci. Lab. Bellevue Paris, vol. 51, 131 (1960).

26. van Steensel, K., Philips Res. Repts., vol. 22, 246 (1967).

27. Hill, R. M., Proc. Royal Soc., vol. 309, 377-417 (1969).

28. Hill, R. M., J. Appl. Phys., vol. 37, 4590-4591 (1966).

29. Hofer, G. and Fromm, E., Phys. Status Solidi (a), vol. 5, 491 (1971).

30. Darmois, E., J. Phys. Radium, vol. 17, 210 (1956).

31. Smythe, W. R., Static and Dynamic Electricity, 2nd ed., McGraw-Hill,

32. See Smythe, reference 31, 118-122.

33. Faraday, M., B. Quaritch, London, 1839, 1855.

34. Stratton, J. A., Electromagnetic Theory, McGraw-Hill, 1941.

35. See Smythe, reference 31, 162.

36. Hobson, E. W., Trans. Camb. Philos. Soc., vol. 18, 277 (1900).

37. Simmons, J. G., J. Appl. Phys., vol. 34, 1793 (1963).

38. Hartman, T. E., J. Appl. Phys., vol. 33, 3427 (1962).

39. Kittel, C., Introduction to Solid State Physics, 3rd ed., John Wiley and Sons, 1968, p. 556.

40. Hill, R. M., Nature, vol. 204, 35 (1964).

41. Henisch, H. K., Rectifying Semiconductor Contacts, Oxford University Press, 1957.

42. Deal, B. E., Snow, E. H., and Mead, C. A., J. Phys. Chem. Solids, vol. 27, 1873 (1966).

43. Saxon, D., and Hunter, R. A., Philips Res. Rept., vol. 4, 81 (1949).

44. Slater, J. C., Insulators, Semiconductors and Metals, vol. 3, McGraw-Hill Book Company, 1967.

45. Ibid., p. 262.

46. Ibid., p. 256-259.

47. Ibid., p. 21.

48. Smith, J. R., Ph.D. Dissertation, Dept. of Phys., The Ohio State University, 1968.

49. Pines, D., Solid State Physics, vol. 1, 374 (1955).

50. Zinan, J. M., J. Phys. Chem., vol. 2, 1704 (1969).

51. Brodsky, M. H., Title, R. S., Weiser, K., and Pettet, G. D., Phys. Rev., vol. B1, 2632 (1970).

52. Cohen, M. H., Fritzsche, H., and Ovshinsky, S. R., Phys. Rev. Lett., vol. 22, 1065 (1969).

53. Mott, N. F., and Gurney, R. W., Electronic Processes in Ionic Crystals, Clarendon Press, Oxford.

54. Davis, K. D., Proc. Conf. on Static Electrification, London, 1967, pp. 29-36.

55. Jonscher, A. K., J. Vac. Sci. Tech., vol. 8, 135 (1971).

56. Friedman, L., Phys. Rev., vol. 135, A1668 (1964).

57. Burger, R. M., and Donovan, R. P., Fundamentals of Silicon Integrated Device Technology, v. 1, Prentice-Hall, Inc., New Jersey, 1967, p. 29.

58. Kronis, H., and Darby, B., Amer. Chem. Soc. Monograph Series, 1922.

59. Harper, R. W., Contact and Frictional Electrification, Oxford at the Clarendon Press, 1967, p. 148.

60. Tamm, I., Phys. Z. Socoj. Um., vol. 1, 733 (1932).

61. Shockley, W., Phys. Rev., vol. 56, 317 (1939).

62. Goodwin, E. T., Proc. Camb. Phil. Soc. Math. Phys. Sci., vol. 35, 221 (1939).

63. Farnsworth, H. E., Schlier, R. E., George, T. H. and Burger, R. M., J. Appl. Phys., vol. 26, 252 (1955).

64. Sherwood, G., J. Appl. Phys., vol. 26, 1238 (1955).

65. Erents, K., and Carter, G., Vacuum, vol. 15, 573 (1965).

66. Holland, L., Brit. J. Appl. Phys., vol. 9, 410 (1958).

67. Schreiner, D. G., Ext. Abstr. 14th AVS Symp. 1966, p. 127.

68. See Burger & Donovan, reference 57, Chapter I.

69. See Harper, reference 59, p. 216.

70. See Harper, reference 59, Chapter XIV.

71. Deal, B. E., and Snow, E. H., J. Phys. Chem. Solids, vol. 27, 1873-1879 (1966).

72. Appel, J., Solid State Physics, vol. 21, 193 (1968).

73. See Maisel, et al., reference 6, p. 1-21.

74. Santeler, D. J., J. Vac. Sci. Tech., vol. 8, No. 1 299-307 (1971).

75. Sauerbrey, G., Phys. Verhandl., 8, 113 (1957).

76. Lostis, M. P., J. Phys. Radium, 20, 25 (1959).

77. Stockbridge, C. D., "Vacuum Microbalance Techniques," vol. 5, 195, Plennar Press, New York (1966).

78. Vacseal is manufactured by the Space Environment Laboratories, Inc., Boulder, Colorado.